case: the culture and sport evidence programme

Understanding the drivers of engagement in culture and sport

Technical Report

July 2010

dcms department for culture, media and sport

ARTS COUNCIL ENGLAND

ENGLISH HERITAGE

MLA MUSEUMS LIBRARIES ARCHIVES COUNCIL

SPORT ENGLAND

The CASE programme The Culture and Sport Evidence (CASE) programme is a three-year joint programme of research led by the Department for Culture, Media and Sport (DCMS) in collaboration with the Arts Council England (ACE), English Heritage (EH), the Museums, Libraries and Archives Council (MLA) and Sport England (SE).

The work on this project was carried out by a consortium led by the EPPI centre with Matrix Knowledge Group

The work in this report was carried out by Matrix Knowledge Group:

Dr Kevin Marsh
Sam MacKay
David Morton
Will Parry
Evelina Bertranou
Dr Rashmi Sarmah

This report can be downloaded from the DCMS website:
www.culture.gov.uk/what_we_do/research_and_statistics/5698.aspx

Contents

1. Introduction

The Culture and Sport Evidence (CASE) programme was set up by the Department for Culture, Media and Sport (DCMS) in 2008, in collaboration with the sector-leading non-departmental public bodies (NDPBs) the Arts Council England (ACE), English Heritage (EH), Museums, Libraries and Archives Council (MLA) and Sport England (SE). The programme aims to generate strategic evidence for *maximising engagement* in culture and sport, and *maximising the value and impacts* people get from engaging in culture and sport. This strategic evidence will be used to inform the deployment of public funds to maximise engagement in sport and culture, and the value citizens in England receive from that engagement.

As part of the CASE programme, DCMS commissioned the EPPI-Centre (Institute of Education, University of London) and the Matrix Knowledge Group to undertake a research project to investigate 'The drivers and value of engagement with culture and sport'. The project used systematic review methods and modelling techniques to begin the process of summarizing existing research evidence on sporting and cultural engagement. This evidence will provide the basis for understanding why people engage in cultural and sporting activities, the value they derive from this engagement and the benefits for society as a whole. An overview of the approach, methods, and results of this project can be found in the '*Understanding the drivers, impact and value of engagement in culture and sport: an over-arching summary of the research*' report published alongside this report on the CASE website.

This report summarises one of three work streams undertaken as part of this project. The objective of this work stream is to answer the question:

What drives engagement in culture and sport?

The other two work streams are concerned with the following research questions:

- What are the **impacts** of engagement?
- What is the **value** of engagement?

The answer to this question has important policy implications as it will directly inform the strategy for meeting the participation targets of the Department for Culture, Media and Sport (DCMS) and associated bodies. That is, a better understanding of the drivers of engagement in culture and sport will ensure policy is focused on the interventions that are most efficient at increasing engagement.

Engagement in culture and sport can take many forms. Thus, before we can consider the drivers of engagement in culture and sport, a more precise definition of engagement is required. This project is concerned with engagement as attendance at culture events / sites and participating in sport. More precisely still, the following definitions are adopted:

- Heritage: attending a heritage site.
- Art: attending an arts event.
- Sport: participating in sport.

- Museums, libraries and archives: attending a museum, library or archive.

Engagement in culture and sport was defined as attendance at culture events / sites and participating in sport as these engagement types are the most common forms of engagement for the different culture and sport sectors. Throughout the remainder of this report, the above engagement types are generically referred to as “engagement in culture and sport”.

The early stages of this project involved a stakeholder engagement exercise to define engagement, and the outcomes of engagement. It is important to note that a number of forms of engagement in culture and sport identified during that exercise are not included in this report, including: deciding, producing / participating in culture, studying, volunteering, and watching sport.

The remainder of this report is divided into four sections. The next section provides a summary of the current understanding of the drivers of engagement in culture and sport. Specifically, it highlights the lessons available from the existing literature in economics and the broader social sciences. A number of themes are identified as consistently emerging from the literature. But it is concluded that attempts to identify the factors that drive demand for culture and sport still face important challenges.

The following two sections present the results of empirical analyses to identify the drivers of engagement in culture and sport. The first method employs regression techniques to analyse existing UK-based survey and administrative data. Acknowledging the limitations of existing survey data, the second method involves the use of system dynamics models to enable more data and more appropriate model structures to be applied to answering the research question.

The final section summarises the report and concludes on the implications for policy making in the area of culture and sport.

2. Background: Our current understanding

What drives engagement in culture and sport? This section summarises the insights generated by attempts to answer this question in the extant literature[1]. It is organised into three sections. The next section highlights the lessons available from the existing economic literature. It is argued that this literature is limited by its focus on conventional econometric approaches informed by the general theory of consumer choice. More precisely, the focus on price and income, at the expense of broader socio-demographic factors, causes econometric approaches to overlook important aspects of the demand for culture and sport. Two such factors emphasised in the culture literature are the cost of the time required to engage in culture and sport, and the idea that preferences for culture and sport are a function of education and previous experience

The following section summarises how the non-economic social science literature can be used to fill some of the gaps in the econometric literature. Even drawing on the insights of this literature, there are, however, still a number of important challenges facing attempts to identify the factors that drive demand for culture and sport. Two approaches to overcoming these challenges are identified. First, more of the same – the application of more sophisticated statistical techniques to more comprehensive data. Second, moving beyond the regression-based approach that has predominated in the literature – acknowledging that, regardless how large the dataset, such approaches might not be able to accurately model the decision to engage in culture and sport in a manner that informs policy.

2.1. The economic literature

Economic approaches to the analysis of the determinants of engagement in culture and sport have, unsurprisingly, tended to start with the general theory of consumer choice in which individuals choose the level of engagement that satisfies their preferences, subject to constraints of income and price (McCarthy et al., 2001). From this perspective, an individual's preferences are assumed to be fixed, dependent on a range of individual characteristics, and generally assumed to be outside of the model (Throsby and Withers, 1979). Consequently, much economic work, both theoretical and empirical, has tended to focus on the effects of income and prices on engagement levels.

Economic theory suggests that as the price of engagement increases, relative to the price of substitute leisure activities, individuals will engage less. Conversely, as income rises and people can afford to engage more, engagement should rise – the income effect. The income effect might, however, be offset by what is referred to as the substitution effect. The substitution effect describes how the opportunity cost of engagement in culture and sport increases as incomes increase, causing the demand for engagement to fall. That is, as wage rates increase, so does what can be earned during leisure time, and so people forgo leisure in favour of more work. The combined impact of the income and substitution effects will

[1] The CASE research database generated as part of this project was drawn on to identify studies included in this section. However, this section does not comprise a systematic review of the studies identified by the EPPI-Centre, as such an undertaking was beyond the scope of this work. The database can be accessed via the CASE website.

depend on an individual's preference for engaging in culture/sport, alternative leisure activities, and his income level (McCarthy et al., 2001).

Empirical work in economics tends to be concerned with estimating the magnitude of these theoretical relationships. That is, to what extent does engagement increase when prices decline (the price elasticity of demand[2]) and when incomes increase (the income elasticity of demand[3])? In the field of culture and sport, much of this empirical work has been undertaken for attendance of the performing arts (Lévy-Garboua and Montmarquette, 2003 – further detail on this literature is included in the remainder of this section). Engagement in culture and sport is often considered a luxury good. This suggests that the demand for engagement would be sensitive to price, falling by relatively large amounts when prices increase. Reviews of the econometric evidence for the attendance of performing arts have, however, tended to confound this expectation (see, for instance, Seaman, 2005). Instead, the literature provides examples of studies that demonstrate price-sensitive demand and price-insensitive demand, as well as mixed results.

Seaman (2005) notes a number of possible explanations for this variation in the estimate of the price sensitivity of demand for the performing arts, including:

1. Distinctions among consumers: Variation in, for instance, income and education could cause variations in how demand responds to changes in price. Specifically, there is evidence that those with higher incomes and education levels might be less price-sensitive.
2. Different pricing strategies: Low sensitivity to prices could be explained by those arts organisations that charge lower than ticket-revenue-maximising prices, as this would cause consumers to be unwilling or unable to aggressively seek substitutes as a reaction to price increases.
3. Variations in the quality of performing arts: People will be less price sensitive with higher-quality performing arts.
4. The expense of engagement: When the opportunity cost of the time required to attend the performing arts is considered, ticket price is only a small element of the expense of attending arts performances.
5. Substitution possibilities: There are few genuine substitutes for the more "esoteric" forms of high art, such as theatre and opera performances (though virtual viewing of such performance on TV would be considered a substitute for some performances), compared with more accessible forms of entertainment, such as cinema.
6. Aggregation of data: Studies that rely on data aggregated across audiences and types of performance tend to demonstrate lower price sensitivities. However, they rely on artificially constructed "average prices". In contrast, price measures in studies of more specific types of engagement are more closely related to what is actually paid, and result in more accurate higher estimates of price sensitivity.

Seaman concludes that further empirical work is required before the effect of price changes on the attendance of the performing arts is understood. In a similar vein, Lévy-Garboua and

[2] Price elasticity of demand is defined as the responsiveness of the quantity demanded of a good or service to a change in its price – the percentage change in demand compared with the percentage change in price. .

[3] Income elasticity of demand is defined as the responsiveness of the quantity demanded of a good or service to a change in income levels.

Montmarquette (2003: 211) note: “It is likely that the demand for the arts is price-elastic and art is a luxury good. But this prediction stems more, as yet, from a theoretical conjecture than from well-replicated empirical estimates.”

Even acknowledging the limitations with econometric analysis of the demand for art, the evidence on the impact of price changes on levels of engagement for other types of culture and sport is less well-developed. Johnson (2003) quotes empirical studies that suggest that the price elasticity of demand for museums is less than one (indicating that demand is not sensitive to changes in prices). Mirroring one of the explanations offered by Seaman (2005) for the low price elasticity of the performing arts, Johnson suggests that this result is a consequence of museums seeking objectives other than profit-maximisation and charging low prices.

There have been some studies of the effects on visitor numbers of museum entrance charges (see, for instance, Anderson, 1998, and Cowell, 2007). These suggest that making entrance to museums free does increase demand. For instance, Cowell (2007) estimates that there were 29 million additional visits to museums in the UK in the five years following the DCMS-sponsored policy of free admissions. He also argues, however, that the data employed to arrive at these estimates is “not quite as robust as they might be” and that further work is required to isolate the precise effect of free admission. For instance, an increase in the number of visits might not translate into the perhaps more desirable effect of increasing the number of visitors. However, it is difficult to distinguish these effects in the current data (Cowell, 2007).

Limited literature was identified on the effect of prices on doing sport. Instead, much of the economic literature on sport has focused on the impact of price on attendance at sporting events, rather than participating in sports (see, for instance, McDonald and Rascher, 2000). A number of studies were identified, however, that examined the relationship between entrance charges to sports facilities and sports participation. Coalter (2002) concluded that cost is only one and mainly not the most important factor explaining non-participation in sport. He concluded that time constraints and geographical proximity play an important role in explaining participation in sport. Kesenne (2006: 76) also argues that participation in sport is relatively insensitive to prices, as “doubling the price for using the [sporting] facility will only reduce the sports activity by 8%, whereas doubling the supply of sports facilities will increase the sports activity by 25%”.

The story is similar for the relationship between income and engagement – the income elasticity of demand for engagement in culture and sport. Once again, much of the relevant literature has focused on attendance at performing arts. Once again, this literature confounds the expectation that attendance at arts performances is a luxury good. If attendance at performing arts is a luxury, a high income elasticity of demand would be observed – a change in income would be associated with a larger change in demand. This is not what is observed in the literature.

In their reviews of the literature, both Lévy-Garboua and Montmarquette (2003) and Seaman (2005) observe a low income elasticity of demand for attendance at performing arts (suggesting that changes in income are not associated with changes in attendance). Lévy-Garboua and Montmarquette (2003: 209) conclude that:

> *Income elasticity estimates are positive, not always statistically significant, and in many studies less than one. This finding, which runs counter to the impression that art goods are luxuries, may be a consequence of the cost of time. Attending live performances is a time-intensive consumption.*

Also emphasising the time-intensive nature of attendance at performing arts, Seaman (2005) argues that the low income elasticity of demand observed in the literature is the result of existing studies not distinguishing between the income effect and the substitution effect. As noted above, these effects tend to offset each other. The income effect describes the positive effect of increased income on demand. The substitution effect describes how the opportunity cost of attendance – the wage rate that could have been earned had attendance at the performing arts event been forgone – increases with income causing demand to reduce with income. The low income elasticity of demand for the performing arts might thus be a consequence of the substitution effect offsetting the income effect.

As with the price elasticity of demand, there is little econometric evidence identified of the income elasticity of demand for attending other types of culture or participation in sport. The exception is Johnson (2003) who quotes studies that suggest that the income elasticity of museum visitor demand is greater than one (suggesting that engagement is influenced by income levels). Following the logic of Seaman's response to the low income elasticity of demand for attendance of the performing arts – that this was the result of the high cost associated with the time required to attend the performing arts – the higher income elasticity of demand for museums could be explained by their requiring less time.

Non-economic studies support the idea that income has a limited impact on participation in sport, and suggest that instead time may be a greater barrier than entrance cost. Lera-Lopez and Rapun-Garate (2007) conclude that personal income is not a barrier to the practice of sport, but that time availability is a major barrier to expand the base of participants or increase the intensity of participation.

2.2 Econometric challenges and the other social sciences

A number of the challenges facing econometric models of the demand for culture and sport can be illuminated by consideration of the empirical evidence generated in the other social sciences. In contrast to economic studies, non-economic empirical assessments of the factors that influence the demand for culture and sport do not adopt a conceptual approach based in economic theory. As noted above, econometric studies operate within the theory of consumer choice, assuming people decide a level of participation that satisfies their (fixed) preferences, subject to the constraints of income and price. While economists tend to treat tastes as being outside of their models, non-econometric studies tend to focus on the role that socio-demographic factors play in influencing individuals' tastes. To this extent, these studies can be viewed as complementing the economic approach by focusing on the socio-economic factors associated with participation as proxies for individual tastes (McCarthy et al., 2001; Gayo-Gal et al., 2006).

The importance of understanding preferences, rather than assuming them to be outside of models, has been highlighted by a number of economists working within the field of culture and sport. On the economic analysis of the demand for art, Lévy-Garboua and Montmarquette (2003: 201) note that:

> *The literature is still groping towards firm answers to simple questions, such as the following: Is art a luxury good? Is it price-elastic or inelastic? Do art goods have close substitutes? However, the consumption of art challenges the conventional assumptions of homogenous goods and services, established tastes, independence of choice among individuals.*

Similarly, when discussing the improvements that could be made to the economic analysis of the demand for the performing arts, Seaman (2005) emphasises the need for a fuller appreciation of the importance of social background variables in reducing the unexplained variation in consumption behaviour. This section considers how insight from the other social sciences can illuminate two of the explanations offered in the previous section for the variation in the findings of econometric studies of the demand for art: the opportunity cost of the time required to attend events; and the impact of the cultivation of taste on demand.

The opportunity cost of time

One possible explanation for the low price and income elasticities of demand for the performing arts is the high opportunity cost of engagement (see the previous section). In support of this argument, Throsby (2001) notes that consumption of culture is time-intensive, suggesting that the price (opportunity cost) of time is likely to be more influential in determining the demand for culture than the ticket price itself (see also Johnson, 2003). Furthermore, other costs of consumption (complements), such as travel, accommodation and meals will also impact on the level of demand. Frey and Meier (2003) note that such complementary costs constitute a high percentage of the total cost of a museum visit, and that the higher these costs the lower the rate of museum visits.

The introduction of the allocation of time into utility theory is generally traced to Becker's Theory of the Allocation of Time (1965). Becker adapted traditional utility theory, which had been concerned with the accumulation of goods subject to the price of goods and income, to allow a trade-off between money income from working and the time spent on non-work activities. As noted above, depending on how an individual's preferences cause them to trade-off income and leisure, an increase in income might cause them to take more leisure time or to work more.

Consistent with the positive (if low) income elasticities of demand for the performing arts noted in the previous section, there is some evidence that the income effect tends to offset the substitution effect. That is, that people will tend to work less as incomes rise. Tisdell (2006) notes that as per capita income levels rise in low income countries, the amount of non-work time tends to rise. Aguiar & Hurst's (2006) research supports this observation, demonstrating that the time spent in leisure in the UK increased from 31 hours per week in 1965 to a high of 37 hours in 1993, before falling to just under 36 in 2003. This observation could imply that more people engage in culture and sport as incomes rise and leisure time

increases. However, Putnam (2000) notes a perceived reduction in leisure time as a result of increased demands on leisure time, or an increased level of competition between leisure activities. Thus, despite an increase in leisure time, the greater variety of types of leisure competing for people's time off could result in a reduced demand for culture and sport. For instance, people might choose to watch more television rather than visit a heritage site or do sport. Haworth and Veal (2004) note that the largest proportion of leisure time is spent watching television.

The importance of time availability for a demand for culture and sport is identified in survey responses that suggest that an actual or perceived lack of time can result in lowered demand (see, for instance, Rowe et al., 2004). Table 1 summarises the results of an analysis of the Taking Part survey demonstrating a positive association between income levels and time being a barrier to engaging in culture and sport

Table 1: Proportion of non-engagers who mention limited time as a barrier to engaging in culture and sport (Taking Part, 2006/7)

	Income level		
	Low	Average	High
Attending arts events	22%	38%	45%
Attending a heritage site	24%	42%	48%
Attending a library	25%	35%	29%
Attending a museum	21%	36%	41%
Attending an archive	14%	20%	17%
Participating in sport	12%	29%	39%

The literature also identifies the impact on engagement in culture and sport of a number of proxies for the availability of time. For instance, having young children has been found to influence engagement in culture (see Bunting et al., 2008; CEBR, 2007; and MLA, 2008). However, this influence is ambiguous, impacting on the demand in a number of competing ways. On the one hand, children will reduce the available time and disposable income to spend on engaging in culture and sport. On the other hand, children act as important drivers of the demand for culture and sport. The Museums, Libraries, and Archives Council (MLA) (2005) note that "the desire for visitors to take their children along was the main reason for attending galleries and museums, with 21% citing this motivation".

Focusing on the importance of time in determining demand for culture and sport allows several important concepts for understanding demand to be introduced into the analysis that have not been adequately dealt with in the empirical literature: the relative preferences for different types of leisure and non-leisure time (McCarthy et al., 2001). The next section considers the development of preferences for culture and sport.

The taste cultivation problem

It is often argued that taste is cumulative, particularly in the area of arts. That is, taste is a function of knowledge, experience and education relative to art, and people who are more knowledgeable about the arts are more likely to attend (McCarthy et al., 2001; Throsby, 2001; Lévy-Garboua and Montmarquette, 2003).

There is much evidence to support this idea within the non-economic literature. For instance, Bunting et al. (2008) analyse the Taking Part survey to identify the socio-demographic factors that impact on arts attendance in England. They conclude that "two of the most important factors in determining whether somebody attends arts activities are education and social status – the higher an individual's level of education and social status, the more likely they are to have high levels of arts attendance" (see also Seaman, 2005; Gayo-Gal et al.., 2006; Sullivan and Katz-Gerro, 2007; and Chan and Goldthorpe, 2005, 2007a, 2007b).

The importance of social status and education is also identified in the analysis of engagement with other types of culture and sport, for instance:

- The Centre for Economic and Business Research (CEBR) (2007) concluded that social status and education had a strong to very strong effect on attendance at heritage sites.
- MLA (2008) identified education as a factor influencing whether people visited either museums or libraries.
- Rowe et al. (2004) identified that participation in sport was significantly skewed towards the professional groups and that sports clubs in England significantly over-represent white, professional males.

Further support for the influence of knowledge and experience of engagement comes from evidence that engagement in culture and sport is influenced by whether a person had the opportunity to engage as a child (for the arts, see, for instance, Oskala et al., 2009). David (2004) reviews the evidence on the role of early learning in influencing participation in sport, concluding that early learning experiences not only develop physical competencies but also the perceptions of competence, both of which are important in the motivation to continued participation. Kay (2004) reviews the evidence on the influence of family on sports participation, concluding that the family is important in introducing children to sport, both as an agent of sports socialisation and as a source of practical support.

These observations have a number of implications for the economic analysis of engagement in culture and sport. As noted by Seaman (2005), lack of consideration of these inter-temporal dynamics might explain the lack of consensus emerging from economic models that conceive of demand as being a function only of income, price, and taste. In response to this challenge, a number of alternatives to the traditional economic model have been proposed. First, Stigler and Becker (1977) suggest that prior experience, knowledge, education, and family background not only influence taste, but allow individuals to become more effective consumers of the arts, causing them to gain more enjoyment from a particular level of consumption.

Second, an institutional perspective deriving influence from Thorstein Veblen's *The Theory of the Leisure Class* (1899) suggests that the satisfaction obtained from engagement derives as much from the instrumental benefits of participation as from direct enjoyment per se (McCarthy et al., 2001). That is, the primary motivation for engagement is to demonstrate social status. This approach is similar to that of sociologist Pierre Bourdieu (1984), who saw culture as being "actively used by members of dominant social classes as means of symbolically demonstrating and confirming their superiority" (Chan and Goldthorpe, 2005: 194).

Chan and Goldthorpe have conducted a number of studies that have tended to reject the notion that the consumption of art is undertaken to demonstrate social status (2005, 2007a, and 2007b). They argue instead for the "omnivore-univore argument". That is, while consumption does map closely to social status, the mapping is not on "elite-to-mass" lines. Instead, the consumption of those of higher social strata includes both "high-brow" and "low-brow" culture – their consumption is not only greater, but also much wider in its range.

Further understanding of the demand for culture and sport requires that these inter-temporal dynamics are incorporated into well-specified demand models (Seaman, 2005). To date, any attempt to capture these effects have tended to be restricted to the inclusion of education or childhood exposure as independent variables. Seaman (2005: 115) summarises the current state of the literature as follows:

> *While many arts demand studies have improved the performance of their estimated equations by including a one-year time lagged dependent variable capturing the effect of past consumption on future consumption, that is not the only, or even necessarily the preferred way to capture the dynamic effect of taste cultivation in arts demand analysis. Even if such a lagged variable is introduced, the underlying theoretical justification for its inclusion is a complex subject, with competing approaches having somewhat different implications.*

2.3 Further challenges and the way forward

The previous sections summarised the current state of the literature on the demand for culture and sport. This identified a number of challenges facing the analysis of demand, including measuring the cost of engagement and incorporating the inter-temporal dynamics of consumption. A range of other challenges also face the analysis of the demand for culture and sport, including:

1. Multi-colinearity: While the conventional wisdom is that education is a key driver of the demand for culture and sport, existing models struggle to confirm this due to the high correlation between education and other variables (Seaman, 2005).
2. Substitutes and complements: There is limited evidence regarding the substitutes and complements of culture and sport, with the prices of neither commonly included in demand models (Seaman, 2005; Lévy-Garboua and Montmarquette, 2003).
3. Quality: Seaman (2005) argues that: "Quality matters, although the mixed results from various objective and subjective measures suggest that we are not yet sure how best to capture this important determinant."

Given the range of challenges facing demand modelling, the remainder of this section considers how research on the drivers of engagement in culture and sport should be taken forward. A key challenge facing demand modelling is missing data. That is, rarely does one good quality dataset contain all the variables required (Seaman, 2005). As Lévy-Garboua and Montmarquette (2003: 209) put it: "The difficulty of gathering good data to account for own price, cross-price, human capital accumulation, learning experience, quality and time cost is obvious. Thus the results are often partial and the methodology varies considerably from one study to another."

Given the challenges of measuring all the factors that influence the demand for culture and sport, it is not surprising that gaps in the current data give rise to calls for the collection of more comprehensive datasets and the application of more sophisticated techniques (see Lévy-Garboua and Montmarquette, 2003). Furthermore, many of these studies pre-date the collection of data such as Taking Part. In this vein, the next chapter reports the results of an exercise to combine and analyse existing survey data to understand the drivers of engagement in culture and sport.

There is also, however, doubt whether the solution to these challenges lies in larger datasets. Seaman (2005) argues that improvements of this type have largely failed to generate substantial improvements in the modelling of the demand for art and instead recommends a greater focus on disaggregated data, clearer definitions of variables such as own price and substitute price, and a better appreciation of the importance of social background variables.

Others make the case for a fundamentally different approach to that which currently dominates the literature. Brook (2005), for instance, argues that there are problems in using survey data to analyse demand as it is unable to capture important variables, such as ease of access to engagement opportunities. Instead, Brook suggests approaches using geographical information systems using data from box offices at theatres and performance arts centres. Applying such a method, Brook found that an index of geographic accessibility and commuting were found to be strongly related to demand.

McCarthy et al. (2001) argue that the prominence of socio-demographic factors in current analysis suffers from two important limitations. First, once a distinction is made between groups of people who participate rarely, those who participate occasionally, and those who participate frequently, socio-demographic variables help to explain little of the variation within these groups. Second, and perhaps most importantly, the focus on socio-demographic variables provides little help to policy makers.

In response to these limitations, McCarthy et al. (2001) suggest that the conceptual approaches that inform current analytical work need improving. That is, current conceptual approaches tend to oversimplify the nature of the decision-making process. Rather than there being one decision, engagement in culture and sport requires a number of sequential decisions, each one influenced by a different set of factors. They illustrate this point as follows:

> *Perhaps the clearest example of why this disaggregation is important is the considerable variation found in the literature on how such economic constraints as ticket prices affect participation behavior. If, as we believe, ticket prices are relevant only for individuals already intending to attend, then estimating how prices will affect participation in the total population (as is implicitly done when a study regards participation as a dichotomous choice) will yield an inaccurate picture of pricing effects.* (31)

The implication of McCarthy et al.'s critique is that the structure of the analysis needs to reflect the structure of the decision problem better. Specifically, the analysis needs to engage with the sequential nature of decisions to engage.

A similar concern for the disregard of econometrics for the “systematic understanding of the structure and operations” of a problem is evident in the system dynamics literature (Arthur, 2008: 21). Proponents of this approach criticise traditional econometric methods for ignoring the structure of underlying models. Sterman (1987) argues that the econometrician’s insistence that model parameters must be estimated with historical data has limited the accuracy of the structure of his models, causing them to rely on complex econometric modelling techniques to try to ensure that the analysis better fits the structure of the problem.

In contrast to econometric approaches, practitioners of system dynamics argue that model specification should not be constrained by the available numerical data, and that there are other sources of data for building confidence in models (Sterman, 1987). As a consequence, system dynamics prioritises the structure of the problem, and emphasises “estimation of parameters outside of the model, using direct observation, interviews of people in the actual system, soliciting expert opinion, participant observation by the modeller, use of focus groups and archival material” (Arthur, 2008: 28). That is, we can use qualitative data to estimate quantitative parameters

Given the willingness to draw on numerous sources of data and emphasise the importance of structural integrity, it is argued that system dynamics is suitable for modelling complex social and economic problems, incorporating complex non-linear systems and causal feedback loops. There are a number of features that distinguish system dynamic approaches, but perhaps the most important in this context is that they “embody rich, operationally realistic, causal theories of how the elements of a complex ‘system’ interact to produce its overall behaviour” (Arthur, 2008: 26).

Given the methodological challenges associated with the current econometric literature – for instance, the limitations of the existing data and the disjuncture between the structure of the analysis and the structure of the decision problem – these two features of system dynamics (its emphasis on problem structure and on drawing on a range of data) suggest that it might offer an alternative way to understand the drivers of engagement in culture and sport. Section 5 of this report presents an exercise to construct a system dynamics model of engagement in culture and sport.

3. The factors that predict engagement: Regression analysis

3.1 Introduction

What drives engagement in culture and sport? The previous section identified a number of challenges faced by extant attempts to answer this question. For instance, existing datasets are rarely comprehensive enough to measure all the variables likely to impact on engagement in culture and sport. This section summarises an attempt to combine existing UK-based survey datasets to overcome this problem.

How do we define engagement in sport and culture? The analysis summarised in this section defines engagement in culture and sport as follows[4]:

1. Heritage attendance– whether a person has visited a heritage site in the past 12 months (linked to the Public Service Agreement (PSA) target for heritage visits by priority groups[5]).
2. Arts attendance – whether a person has attended an arts event in the past 12 months (participating in arts what not included in the analysis.
3. Sports participation – whether a person has done three episodes of at least 30 minutes of moderate-intensity sporting activity in the past four weeks (as defined in the Sport England "1 million" indicator).
4. Library attendance – whether a person has visited a library in the past 12 months.
5. Museum attendance – whether a person has visited a museum in the past 12 months.

The next section describes the data sources and methodological framework employed in analysis. The following five sections present the specific method and results for each of the culture/sport types listed above. The final section summarises the analysis findings, the implications for policy, and how the analysis has been able to overcome the challenges faced by similar analysis in the extant literature.

3.2 Data sources

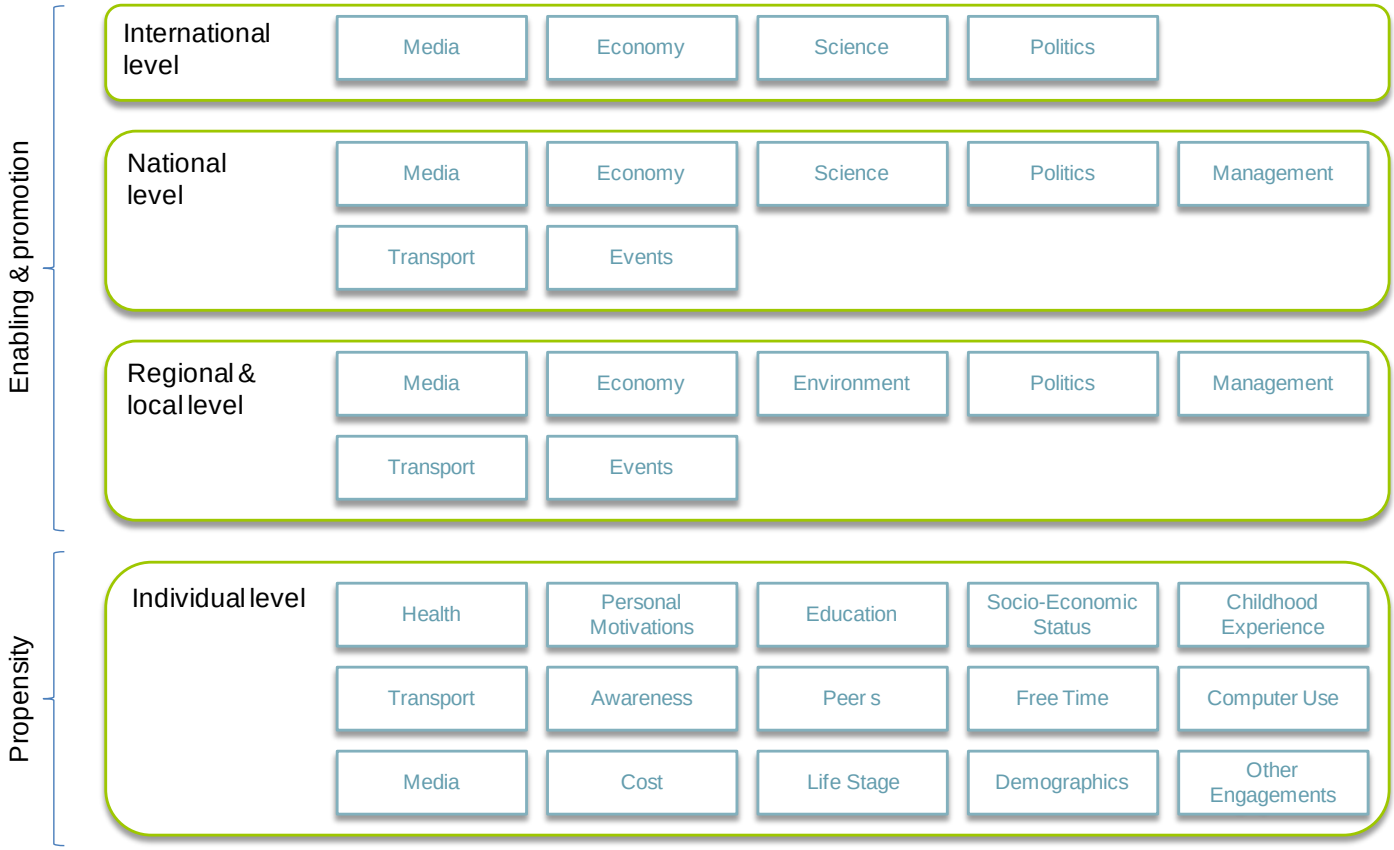
A rapid review of the existing literature in the art and sport sectors, including that referred to in section 3, was undertaken to develop a conceptual framework of the drivers of engagement in culture and sport. Both the theoretical and empirical relationships between factors and engagement were identified in the literature, and organised into a framework. Figure 1 summarises final framework, which identifies the driver-types that influence engagement. Further detail on the review and the conceptual drivers identified are available

[4] The analysis did not cover attending an archive as only a small proportion of the population engaged in this activity, reducing the likelihood that any analysis would be able to explain variation in attendance. For further detail on the rationale behind this choice of engagement definition, see the introduction to this paper.

[5] People with a physical or mental disability, from black or minority ethnic groups or in certain National Readership Survey social grade groups (C2, D and E).

in Appendix 1. This framework was used to assess the availability of data in different survey and administrative sources with which to construct the regression analysis.

Figure 1: Conceptual drivers of engagement in culture and sport



A wide range of data sources were rapidly assessed in terms of their methodological quality, geographic definition, available time points, and their ability to measure the conceptual drivers identified in the literature. The purpose of this assessment was to identify a main data source with which to model engagement in culture and sport, and any potential data sources that could be efficiently exploited to supplement the main source. Appendix 2 contains further detail on the data sources reviewed.

The data source selected to provide the majority of the variables in the analysis was the adult Taking Part 2007/08 survey. It was chosen because it was recent, methodologically robust and contained all of the engagement types of interest and a very rich set of explanatory variables. This dataset was then supplemented with variables from area-level datasets in order to populate the conceptual framework as comprehensively as possible. The following data were used to populate the models:

1. Taking Part 2007/08 survey data – for individual-level engagement and explanatory variables.
2. Asset data – for area-level counts of assets (facilities, organisations, etc.) by local authority, sourced from the Culture and Sport Evidence programme (CASE) member organisations and Experian.

3. National Indicator data – for area-level estimates of factors such as asset quality, community cohesion and the index of access to public services by local authority. These data are compiled by local authorities and include estimates calculated from the Place Survey.

None of the datasets used provided data to measure the cost of engaging. Further information on the data used in the models is available in Appendix 3.

3.3 Attending an art event

3.3.1 Model development

The initial art model was built using the response variable **art** – whether the respondent had attended an arts event in the past year. The model was constructed by populating the conceptual framework summarised in Appendix 1. The variables used to construct this model are described in Table *2*. More information on these variables is available in Appendices 3 and 4.

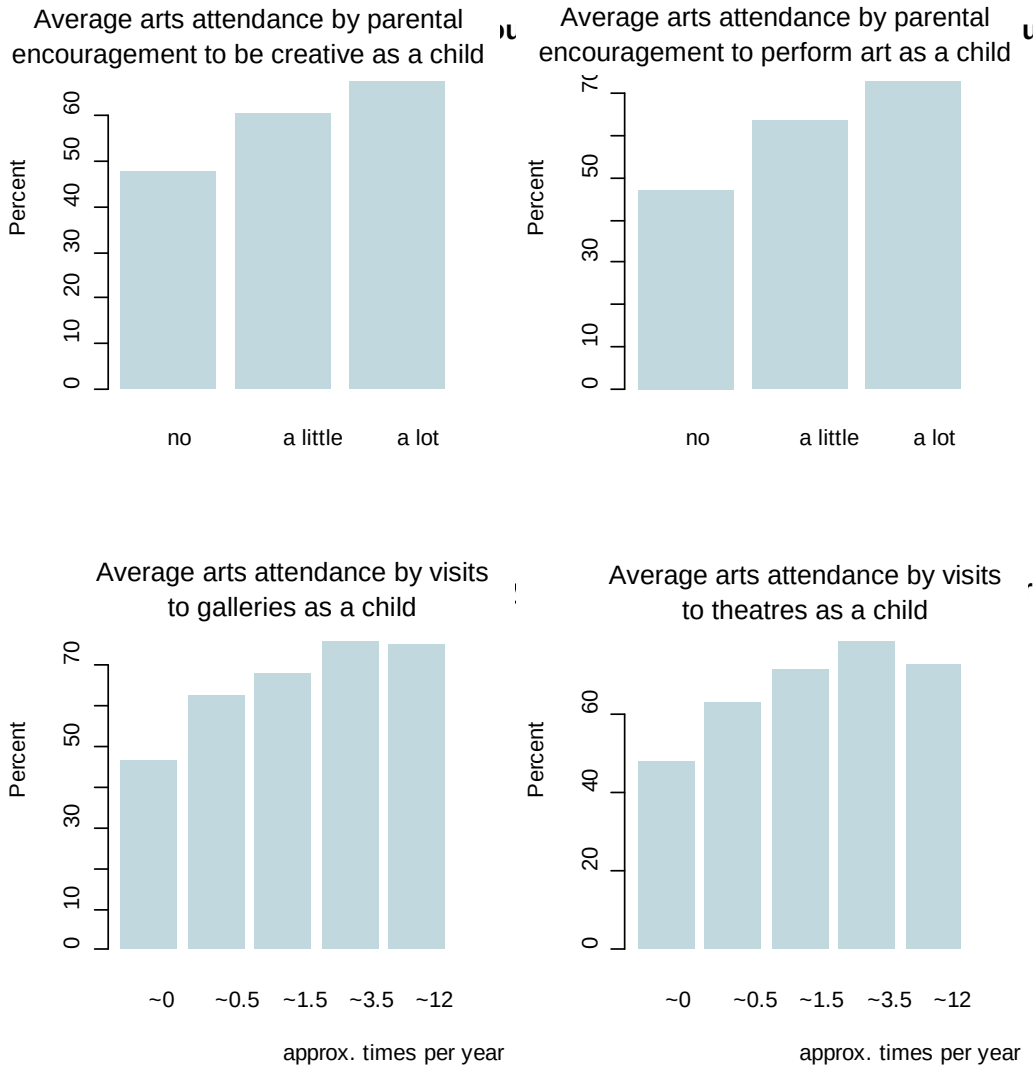
Table 2: Variables used to build art model

Variable name	Description
access	Accessibility to services index from National Indicator set (percent) – percentage of people who can reach selected core services and facilities within 15 minutes by public transport and/or walking (National Indicator 175)
age	Age in years
alcohol	Days alcohol drunk per week (categorical from 0=none to 4=every day)
art *(response variable)*	Attended arts event in last 12 months (1=yes, 0=no)
art_internet	Visited art website in last 12 months (1=yes, 0=no)
art_tv	Watches art television programmes (1=yes, 0=no)
BMEgroup	Member of ethnic minority group (1=yes, 0=no)
child_encouraged_creative	Encouraged to draw, paint or write when growing up (categorical from 0=not to 2=a lot)
child_encouraged_perform	Encouraged to perform music, dance, act growing up (categorical from 0=not to 2=a lot)
child_gallery_visit	Average times per year taken to museums or art galleries when growing up
child_theatre_visit	Average times per year taken to theatre, dance event or classical music when growing up
children	Children living in household (1=yes, 0=no)
community_cohesion	% of people who feel that they belong to their neighbourhood
coupled	Living as part of a couple (1=yes, 0=no)
cultural_influence	Has influence over local cultural facilities (categorical from 0=no to 2=a lot)
cycles	Cycles to get from place to place (1=yes, 0=no)
education	Highest educational qualification held (categorical from 1=degree to 8=no qualifications)
highSES	Member of high socio-economic group NS-SEC 1-4 (1=yes, 0=no)
income	Highest income in household (interval from 0=£0 to 12=£50,000 or more)
internet	Has access to internet (1=yes, 0=no)
limiting_illness	Has illness or disability which limits activities (1=yes, 0=no)
local_art_awards	Population weighted count of Artsmark 'Regularly Funded Organisations' awards in area

Variable name	Description
local_art_funding	Population weighted sum of Artsmark 'Regularly Funded Organisations' funding in area
local_gallery_satisfaction	% of people who are very or fairly satisfied with museums/galleries
motor_vehicle	Has access to a motor vehicle (1=yes, 0=no)
newspaper	Reads daily newspaper at least three times per week (1=yes, 0=no)
radio	Radio available in household (1=yes, 0=no)
religious	Religiosity (0=not religious, 1=non-practising, 2=practising)
sex	Sex of respondent (1=male, 0=female)
social_housing	Is a social housing tenant (1=yes, 0=no)
tv_hours	Hours of television watched per day on average
work_status	Work status (0=not working, 1=part-time, 2=full-time)

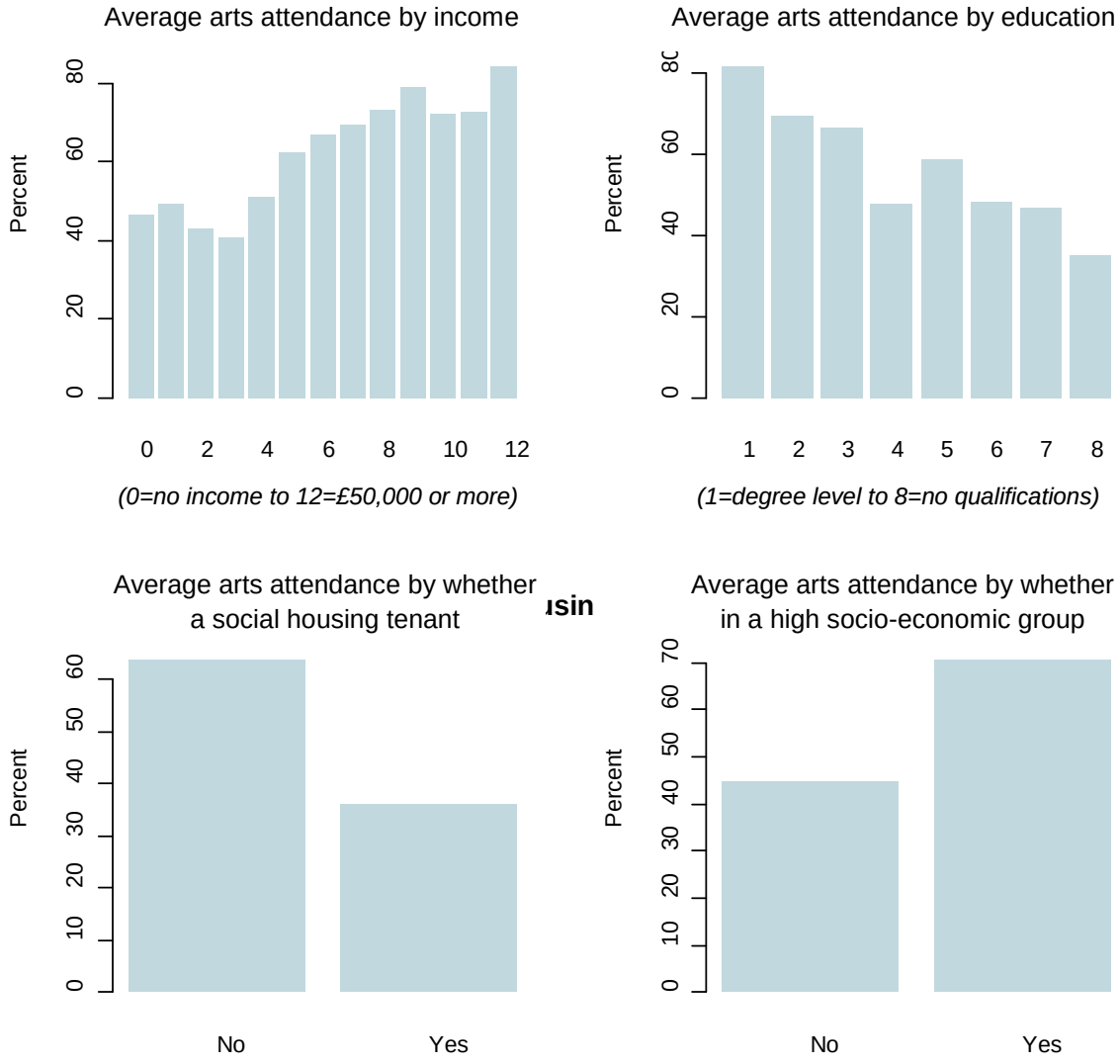
Bivariate analyses of the explanatory variables with **art** identified many significant associations (see Appendix 5). Some of the strongest associations involved education, socio-economic, media and childhood experience factors. *Figure 2* shows plots demonstrating the consistently positive bivariate association between childhood experience measures and art attendance.

Figure 2: Proportion attending arts events by childhood experience



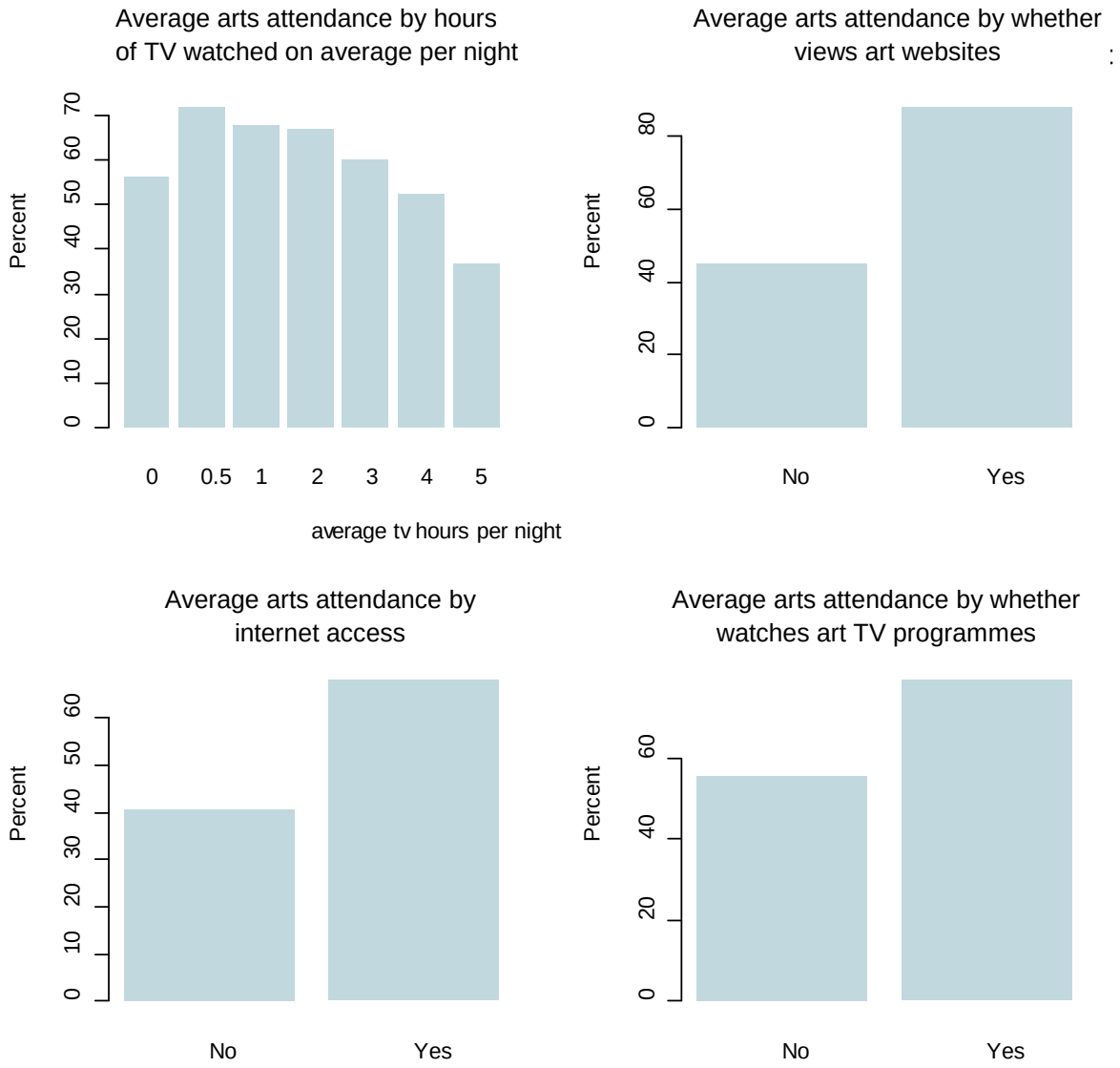
Measures related to socio-economic status and education exhibited positive bivariate associations with art attendance. Attendance was positively associated with income, education and socio-economic status, and negatively associated with social housing. Figure *3* shows plots of art attendance by income, education, social housing and socio-economic status.

Figure 3: Proportion attending an art event by socio-economic and education characteristics



Measures of media consumption also exhibited positive and negative associations with art attendance. Although watching art television programmes was positively associated with attendance, increasing hours of average television viewing was negatively associated with attendance. Internet use, however, was positively associated with attendance, both when measured as internet access and viewing art-related websites. Figure *4* shows plots of these associations.

Figure 4: Proportion attending an art event by television and internet consumption measures



Interval and ratio variables were plotted to determine whether any exhibited a high degree of skew, which could lead to instability in model estimation (see Appendix 8). The original variables **local_art_awards** and **local_art_funding** were intended as measures of supply and were simple counts of Arts Council awards and funding by local authority – they did not take account of the population in those authorities. In order to take account of local authority population size and so produce a measure of supply density, the data were weighted by 2007 mid-year ONS population estimates (in thousands). These variables were highly skewed and so they were log transformed ($\log_e(X+0.01)$) in order to reduce skew.

The supply variables are closely related to each other, with organisations that receive an award also receiving funding. An analysis of the correlation between the log transformed variables confirmed this (R=0.81). In order to ensure that multicolinearity would not cause

problems during model selection, only **local_art_awards** was retained in the initial model to avoid any estimation issues from multicolinearity between the two variables.

Previous research on the drivers of engagement has suggested that multicolinearity among explanatory variables (e.g. **income** and **education**) leads to difficulties in isolating their effects on engagement (see section 3). In order to test whether multicolinearity was likely to be a problem in our data, variance inflation factors (VIF) were calculated (see Appendix 6). None of the VIFs indicated that problematic multicolinearity was present in the explanatory variables (based on a rule of thumb that VIFs should not exceed 10).

The multivariate model constructed was of the form:

$$\begin{aligned} L_{ij} &= Logit\left\{\Pr(y_{ij}=1 \middle| x_{1j}, x_{2j}, x_{3ij}, x_{4ij}, ..., x_{nij}, \zeta_j)\right\} \\ &= \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + ... + \beta_n x_{nij} + \zeta_j \end{aligned} \quad (1)$$

Where:

$i =$	Identifier for individual
$j =$	Identifier for area
$L_{ij} =$	Natural logarithm of odds of 'yes' response to a yes/no question
$y_{ij} =$	Response to yes/no question as a Boolean value
$\beta_0, \beta_1, \beta_2 ... =$	Coefficients of regression model
$x_{1j}, x_{2j} ... =$	Explanatory variables at area level
$x_{3ij}, x_{4ij}, ..., x_{nij} =$	Explanatory variables at individual level
$\zeta_j =$	Random intercept for area

The majority of the data used to construct the models were at the individual level. Due to the difficulty in including all relevant variables at the area level, the model specification included a random intercept term. This term increases the flexibility of the logistic model to give a separate intercept for each Local Authority, thereby allowing for heterogeneity in engagement levels due to unmeasured area level factors and increasing the accuracy of estimates.

The data were weighted using the variables **pweight1** (individual level weight) and **pweight2** (area level weight, see Appendix 7 for more information). Model selection proceeded by removing terms in order of decreasing significance. A nominal significance level of 5% was used to determine whether terms were retained in the model. Once the terms remaining were all significant at the 5% level, interactions to identify additional variations in effect due to demographics, family structure, socio-economic group and free time were added.

At this point, the effect of viewing art websites (**art_internet**) was identified as problematic. The effect size was much greater than for any other covariate. The likely reason for this very large effect was ambiguity in the causal direction, as visiting art websites might increase the

likelihood of attending arts events, but attending arts events can also increase the likelihood of visiting art websites. Therefore, **art_internet** was removed from the model, and some previously borderline significant terms were retested. More details of the model selection process and Stata code used can be found in Appendix 7.

3.3.2 Model results

Table 3 shows the final estimates from the multivariate modelling of art attendance. The bar chart at the side of the table demonstrates the strength of relationships – red bars for negative effects and green bars for positive effects. The bar sizes are proportional to either the odds ratio (where a variable is binary) or the odds ratio raised to the power of one standard deviation (where a variable is continuous or interval), demonstrating the relative effect of changing a binary variable from 0 to 1, or moving one standard deviation along the scale of a continuous variable

The estimates reflected a similar picture to the bivariate analyses, with moderate to large positive effects being in evidence for variables measuring media consumption, socio-economic indicators, education and childhood experience. Additionally, however, large effects were in evidence for cultural influence, alcohol consumption, age and being part of a BME group.

Table 3: Multivariate analysis of the factors associated with attendence at arts events

Model terms	Odds ratio	Coefficient	Std. Err.	p-value	Less/More likely
Socio-economic characteristics					
Member of a BME group vs. not	1.7565	0.5633	0.2957	0.057	
Member of a BME group interacted with age	0.9799	-0.0203	0.0066	0.002	
Age	1.0434	0.0425	0.0102	0.000	
Age squared	0.9996	-0.0004	0.0001	0.000	
Religion: non-practicing vs. not religious	0.8554	-0.1562	0.0986	0.113	
Religion: practicing vs. not religious	1.1647	0.1524	0.1011	0.132	
High social-economic status vs. not	1.3188	0.2767	0.0715	0.000	
Income	0.9857	-0.0144	0.0386	0.709	
Income squared	1.0061	0.0061	0.0030	0.039	
Alcohol drunk 1 day/week vs. none	1.4813	0.3929	0.0853	0.000	
Alcohol drunk 2 days/week vs. none	1.9695	0.6778	0.0963	0.000	
Alcohol drunk 3 days/week vs. none	2.0211	0.7036	0.1360	0.000	
Alcohol drunk 4 days or more/week vs. none	1.4941	0.4015	0.1280	0.002	
Sex: male vs. female	0.6129	-0.4896	0.1049	0.000	
Sex interacted with living as a couple	1.3271	0.2830	0.1267	0.025	

Factors that predict engagement: regression analysis

Model terms	Odds ratio	Coefficient	Std. Err.	p-value	Less/More likely
Living as a couple vs. not	0.8952	-0.1107	0.0970	0.254	
Living in a couple interacted with children	0.7394	-0.3020	0.1519	0.047	
Children living in household vs. not	0.8077	-0.2135	0.1331	0.109	
Work status: part time vs. not working	1.0819	0.0787	0.1097	0.473	
Work status: full time vs. not working	0.8100	-0.2107	0.0986	0.033	
Social housing tenant vs. not	0.6727	-0.3965	0.0913	0.000	
Educated to other higher level vs. degree	0.5539	-0.5907	0.1214	0.000	
Educated to A-level vs. degree	0.7160	-0.3341	0.1174	0.004	
Educated to trade apprentice level vs. degree	0.4273	-0.8502	0.1587	0.000	
Educated to 5+ GCSEs A*-C level vs degree	0.5455	-0.6061	0.1139	0.000	
Educated to <5 GCSEs A*-C level vs degree	0.4253	-0.8548	0.1592	0.000	
Educated other qualifications vs degree	0.3901	-0.9415	0.1850	0.000	
Educated to no qualifications vs. degree	0.3601	-1.0214	0.1231	0.000	
Illness or disability vs. not	0.7805	-0.2479	0.0832	0.003	
Media access					
Watches art on TV vs. not	2.2932	0.8299	0.1049	0.000	
Radio available in house vs. not	1.6345	0.4913	0.1430	0.001	
Has access to internet vs. not	1.5626	0.4464	0.0807	0.000	
Hours of TV watched per day	0.8877	-0.1191	0.0263	0.000	
Childhood experience					
Encouraged to perform art as child: some vs. not	1.3382	0.2913	0.0863	0.001	
Encouraged to perform art as child: a lot vs. not	1.5913	0.4646	0.0954	0.000	
Encouraged to paint or draw as child: some vs. not	1.1056	0.1004	0.0946	0.289	
Encouraged to paint or draw as child: a lot vs. not	1.3271	0.2830	0.0828	0.001	
No. of times taken to the theatre p.a. as child	1.1109	0.1051	0.0188	0.000	
No. of times taken to a gallery p.a. as child	1.0344	0.0338	0.0175	0.053	
Influence over provision					
Influence over cultural facilities: some vs. none	1.6785	0.5179	0.1080	0.000	
Influence over cultural facilities: lot vs. none	1.9297	0.6574	0.3768	0.081	
constant	-	*-1.0155*	*0.3052*	*0.001*	

number of individual respondents = 9858

number of local authorities = 346

variance of random intercept = 0.1933

Model terms	Odds ratio	Coefficient	Std. Err.	p-value	Less/More likely
standard error of random intercept = 0.0457					

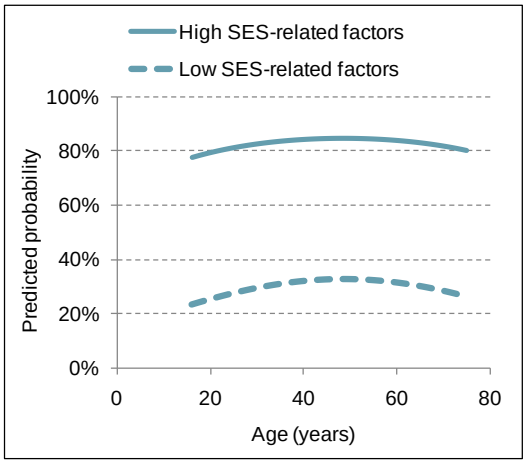
(Note: the bar chart is approximate - the sizes of bars are proportional to the odds ratios raised to the power of one for dummy variables or the odds ratios raised to the power of one standard deviation for continuous variables)

Predictions were created using the model for various combinations of the explanatory variables.

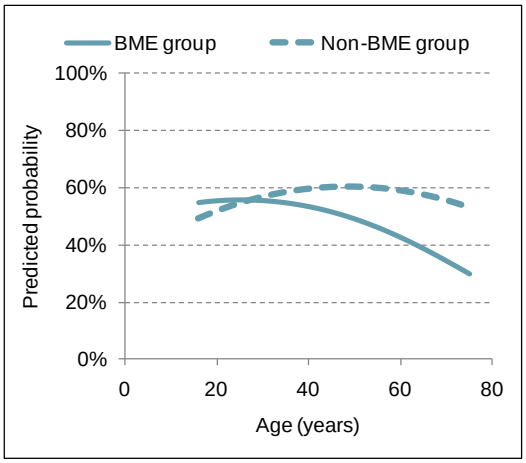
Figure 5 shows a comparison of the predicted probability of attendance at arts events by differences in socio-demographic characteristics. *Figure 5a* shows a comparison of the predicted probability of attendance at arts events by age for individuals high on all factors related to socio-economic advantage with those who are low on the same factors. The dashed line represents individuals who are of low socio-economic status, have an annual income of £10,000 to £14,999, who watch four hours of television per night, are social housing tenants and have an education level of fewer than five A*-C grades at GCSE. The solid line represents individuals who are of high socio-economic status, have an annual income of £50,000 or more, who watch two hours of television per night, are not social housing tenants and have an education level of degree or above. Combined together, these factors have a very large impact on the predicted probability of engaging with art, with 80% of high socio-economic individuals predicted to engage compared with only 30% of low socio-economic individuals.

Figure 5: Predicted probability of attendance at art events by groups of factors

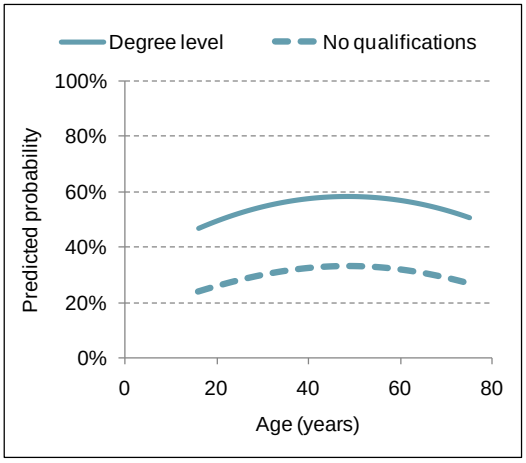
5a. High and low SES-related factors



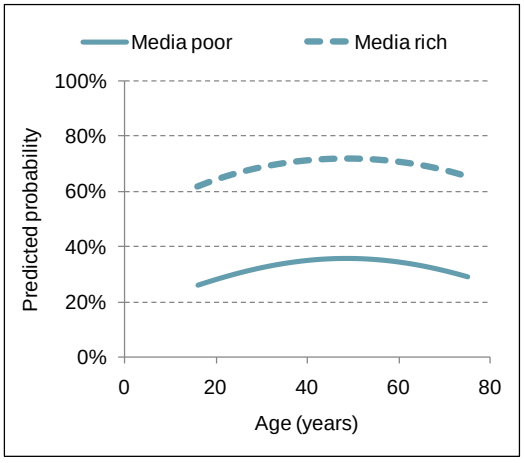
5b. BME and non-BME groups



5c. Qualification levels



5d. Media consumption



5e. Childhood experience

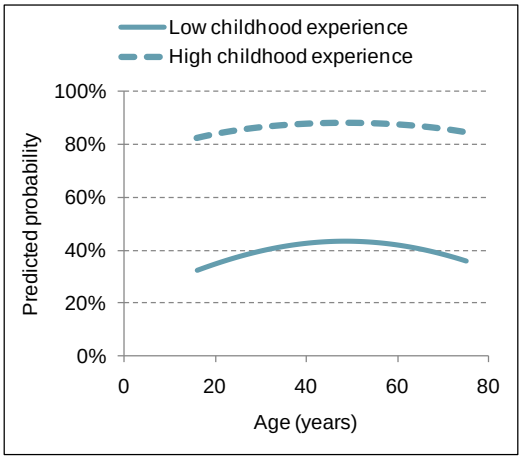


Figure 5b shows a comparison of the predicted probability of attendance at art events by age for individuals in a BME and non-BME groups. The solid line represents individuals who are in a BME group, the dashed line represents individuals who are not. Interestingly, young BME individuals are predicted to be more likely to engage with art than non-BME, but this predicted engagement declines steeply with age, so that at around 30 years of age BME engagement is lower than for the equivalent non-BME individuals.

Figure 5c shows the effect of education alone on predicted attendance at art events. The solid line represents individuals with degree-level education, while the dashed line represents the same individuals but with no qualifications at all. The difference in levels of engagement are predicted to be approximately 20% across the age range, however it is likely in practice that this difference in education level would be accompanied by other differences, such those associated with socio-economic factors in *Figure 5a*.

Media consumption was also shown to be strongly associated with engagement with art. *Figure 5d* shows a comparison of the predicted probability of attendance at art events by age for individuals who are media rich with those who are media poor. The dashed line represents individuals who have access to the internet, watch art tv programmes, watch 2 hours of television per day and have a radio. The solid line represents individuals who watch no television and do not have access to the internet or a radio. Combined together, these factors have a large impact on the predicted probability of engaging with art, with 70% of media rich individuals predicted to engage compared to 30% of media poor individuals.

Some of the most dramatic effects related to childhood experience of engagement with art. *Figure 5e* shows a comparison between predicted attendance at arts events for individuals who as children were taken to art galleries and arts performances once a month on average, were encouraged “a lot” to be artistically creative and encouraged “a lot” to perform art (dashed line) with individuals who, as children, were taken to art galleries and arts performances only once a year on average, were not encouraged to be artistically creative and were not encouraged to perform art (solid line). The average level of engagement predicted for individuals with low childhood experience of art is approximately 40%, whereas for high childhood experience it is almost 90%.

3.3.3 Model testing and diagnostics

Software for multilevel modelling is relatively new and developing as computers become more powerful and are able to cope with the large number of calculations needed to estimate these types of model. As a result, the number of fit and diagnostic tests available for use with weighted multilevel models is limited. In order to test the final art model thoroughly, a single-level logistic regression model with robust standard errors was constructed to mimic the multilevel model. Apart from not containing a random intercept parameter, this model contained the same terms as the final art model. In the model, the random intercept term was significant at the 5% level, thus a single-level model does not fit the data as well as the multilevel model. Therefore, if the single-level model was found to fit the data well, using the tests available for single-level unweighted models, it is reasonable to assume that the multilevel model would also fit the data well.

A model specification link test of the single-level model did not identify any problems with the model specification (p=0.172) and a Hosmer-Lemeshow test also suggested that the model was a good fit to the data (50 groups, $Chi^2(48)=58.74$, p=0.1378). Diagnostic plots were produced using the multilevel art model. A plot of standardised residuals of the local authority level is shown in *Figure 6*, and a plot of deviance residuals for the respondent level is shown in Figure *7*. The plots do not demonstrate any abnormalities to cause concern. Further information on model testing and diagnostics can be found in Appendix 7.

Figure 6: Local authority level standardised residuals

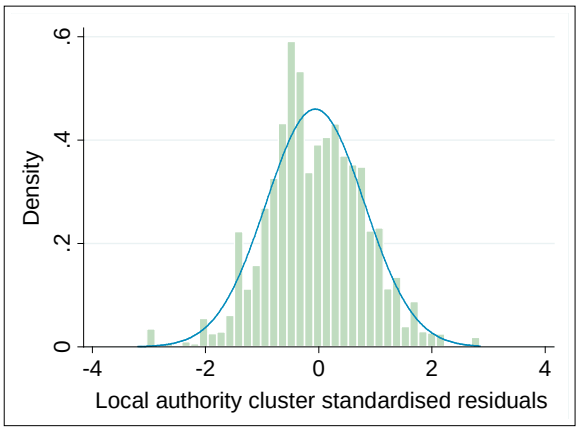
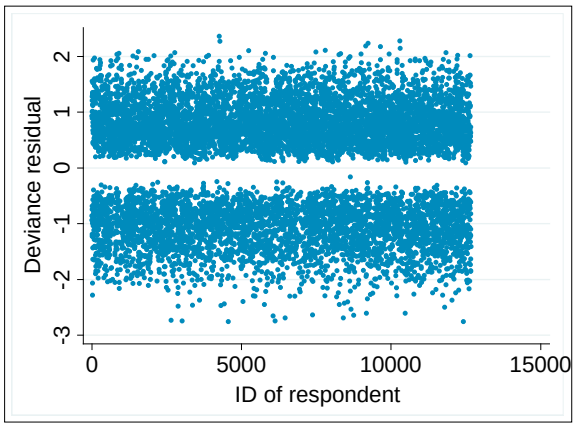


Figure 7: Respondent level deviance residuals



Estimates of explanatory power for logistic regression (known as pseudo-R^2 measures) do not benefit from the straightforward interpretation found in continuous regression models because the response variables can only take the values 0 or 1. As a result, although many different versions are available, statisticians do not view any of these measures as particularly robust and they only serve as weak analogues of the more familiar R^2 in continuous regression. Generally, pseudo-R2 measures return low values of explanatory power of between 10% and 30%. Despite these reservations, two estimates were calculated using the single level robust weighted model to give comparable indications of explanatory power:

- McFadden's R^2 – this resulted in an estimate of a 19.3% reduction in error using the model to predict engagement as compared to the overall mean.
- McKelvey and Zavoina's R^2 – this resulted in an estimate of 33.3% reduction in error using the model to predict engagement as compared to the overall mean.

3.4 Visiting a heritage site

3.4.1 Model development

The heritage model was built using the response variable **heritage** – whether the respondent had visited a heritage site in the past year[6]. The model was constructed by populating the conceptual framework (Appendix 1). The variables used to construct this first model are described in *Table 4*. More information on the data is available in Appendices 3 and 4.

Table 4: Variables used to build heritage model

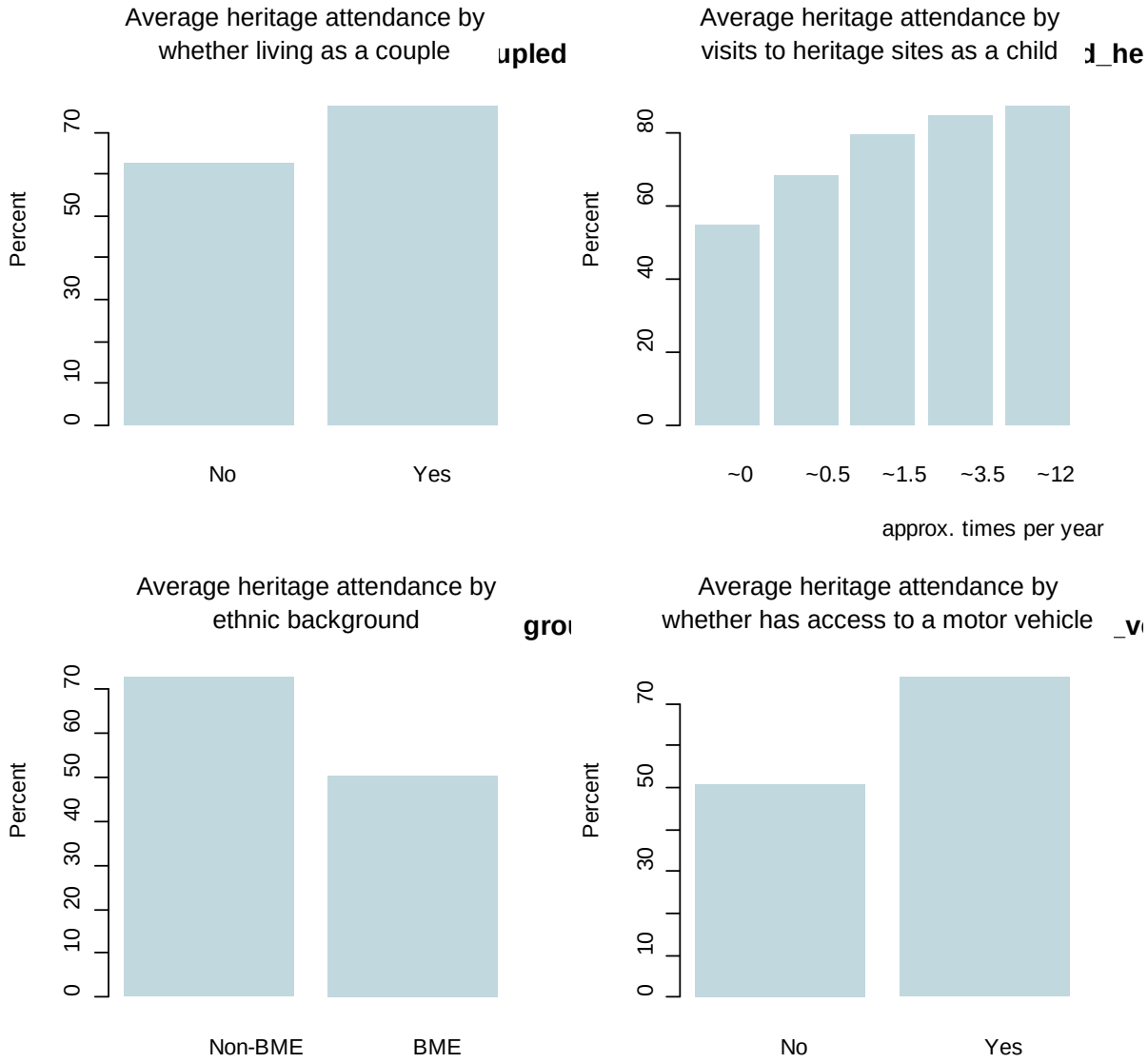
Variable name	Description
access	Accessibility to services index from National Indicator set (percent) – percentage of people who can reach selected core services and facilities within 15 minutes by public transport and/or walking (National Indicator 175)
age	Age in years
alcohol	Days alcohol drunk per week (categorical from 0=none to 4=every day)
BMEgroup	Member of ethnic minority group (1=yes, 0=no)
child_heritage_visit	Average times per year taken to heritage sites when growing up
children	Children living in household (1=yes, 0=no)
community_cohesion	% of people who feel that they belong to their neighbourhood
coupled	Living as part of a couple (1=yes, 0=no)
cycles	Cycles to get from place to place (1=yes, 0=no)
education	Highest educational qualification held (categorical from 1=degree to 8=no qualifications)
heritage *(response var.)*	Visited heritage site in past 12 months (1=yes, 0=no)
heritage_internet	Visited heritage website in past 12 months (1=yes, 0=no)
highSES	Member of high socio-economic group NS-SEC 1-4 (1=yes, 0=no)
history_tv	Watches historical television programmes (1=yes, 0=no)
income	Highest income in household (interval from 0=£0 to 12=£50,000 or more)
internet	Has access to internet (1=yes, 0=no)
limiting_illness	Has illness or disability which limits activities (1=yes, 0=no)
local_heritage_sites	Population weighted count of heritage sites in area (including scheduled monuments and grade I and II* listed buildings)
motor_vehicle	Has access to a motor vehicle (1=yes, 0=no)
newspaper	Reads daily newspaper at least three times per week (1=yes, 0=no)
radio	Radio available in household (1=yes, 0=no)
religious	Religiosity (0=not religious, 1=non-practising, 2=practising)
sex	Sex of respondent (1=male, 0=female)
social_housing	Is a social housing tenant (1=yes, 0=no)

[6] At the request of English Heritage, an additional model was developed of heritage engagement defined as whether a person has visited a *non-religious historic building open to the public* in the last 12 months. Model selection was undertaken using robust single level logistic regression, with the final model estimation including a random intercept. The results of this additional analysis can be found in Appendix 16.

Variable name	Description
tv_hours	Hours of television watched per day on average
work_status	Work status (0=not working, 1=part-time, 2=full-time)

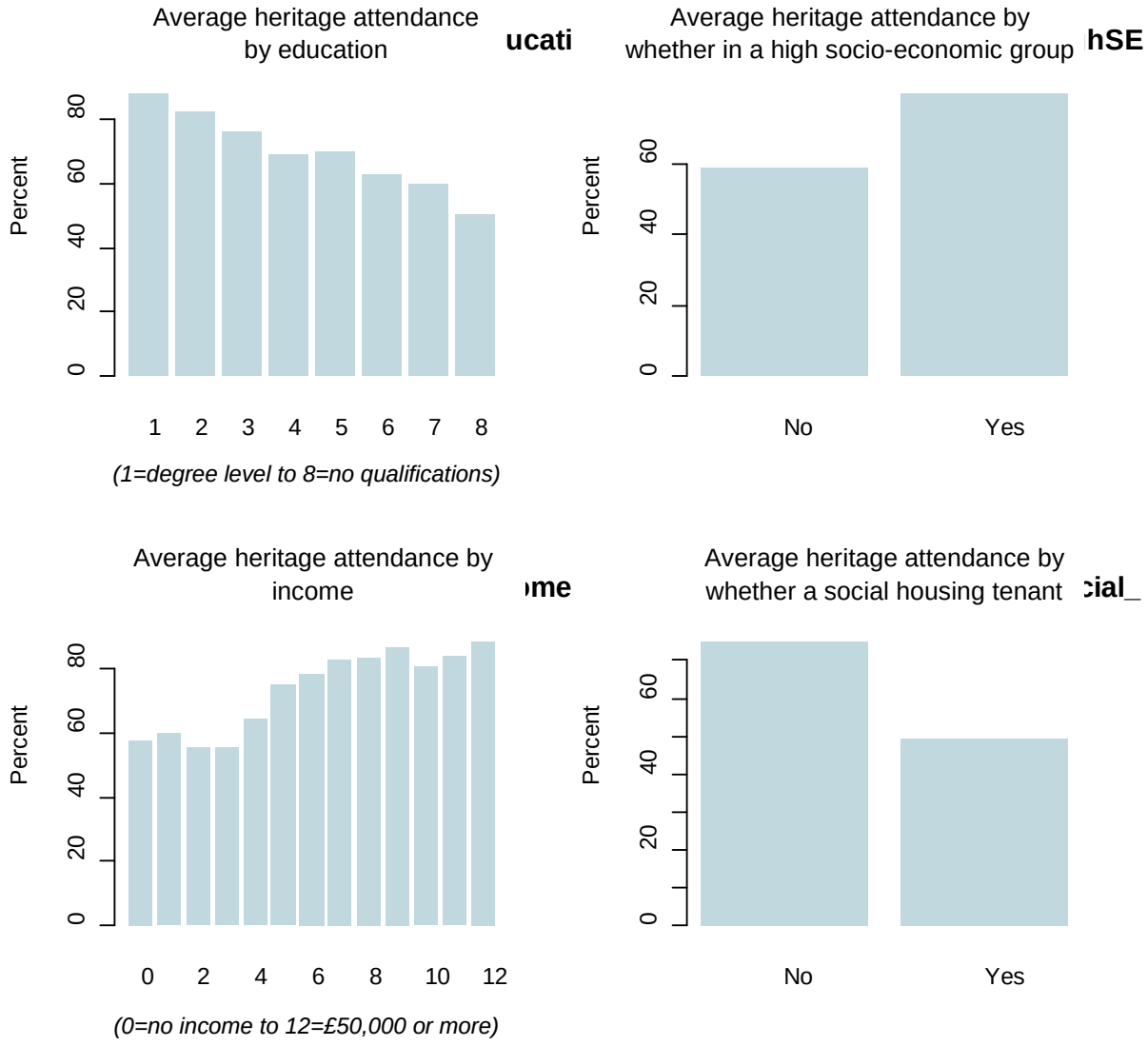
Bivariate analyses of the explanatory variables with **heritage** identified many significant associations (see Appendix 5). Some of the strongest associations involved childhood experience, ethnicity, motor vehicle access, socio-economic and media factors. *Figure 8* shows plots of visiting heritages sites against couple status, childhood experience of visiting heritage sites, ethnicity and access to a motor vehicle.

Figure 8: Proportion visiting a heritage site by couple status, childhood experience, BME group and access to a motor vehicle



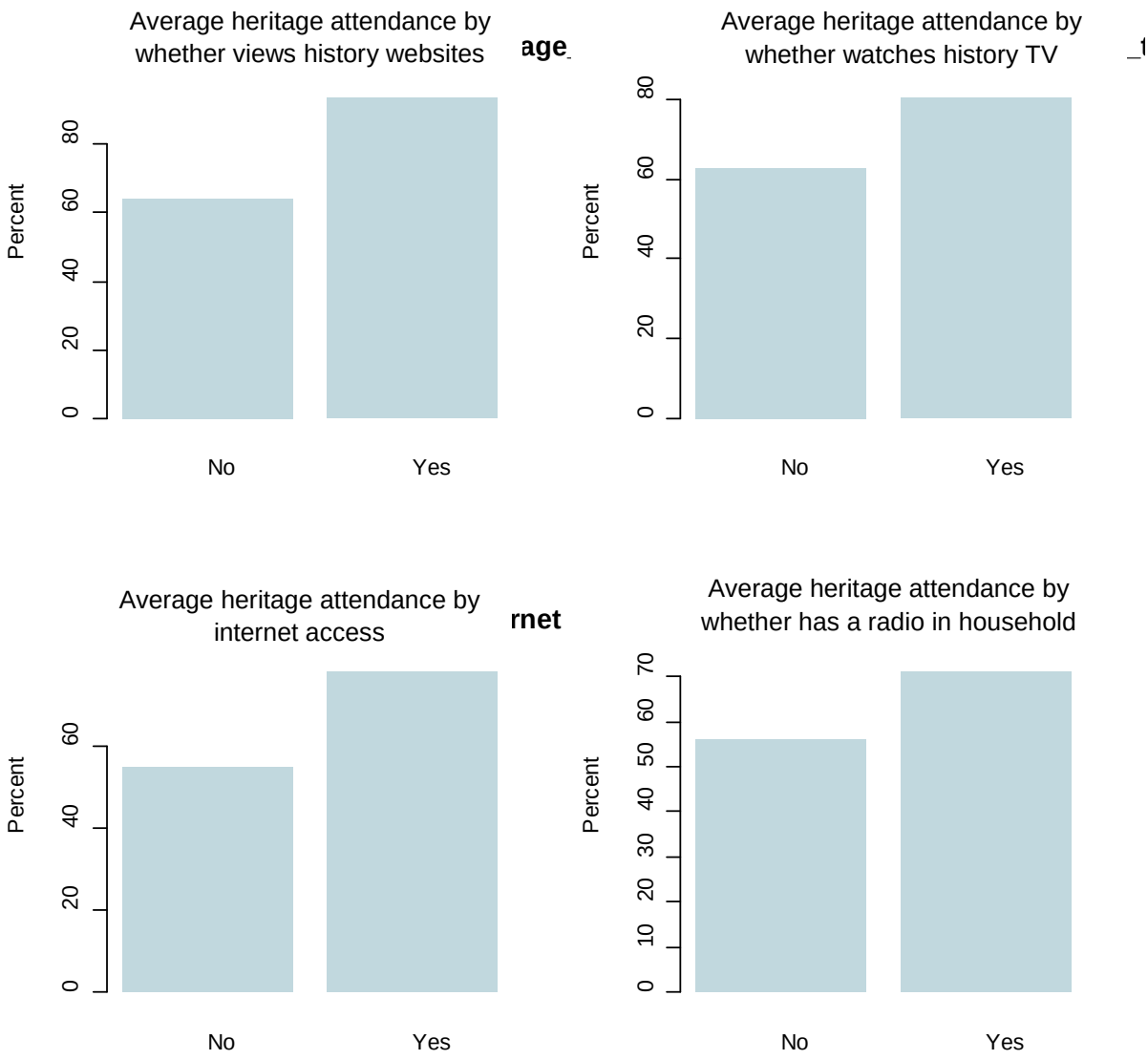
Measures related to socio-economic status and education exhibited positive and negative bivariate associations with visiting a heritage site. The probability of visiting was positively associated with income and socio-economic status, and negatively associated with lower education levels and social housing. *Figure 9* shows plots of heritage engagement by income, education, social housing and socio-economic status.

Figure 9: Proportion visiting a heritage site by socio-economic and education characteristics



Measures of media consumption exhibited strong positive associations with visiting a heritage site. Visiting heritage websites, watching history TV programmes, internet and radio access were all associated with an increased probability of visiting a heritage site. *Figure 10* shows plots of these associations.

Figure 10: Proportion visiting a heritage site by internet, TV and radio consumption measures



Interval and ratio variables were plotted to determine whether any exhibited a high degree of skew, which could lead to instability in model estimation (see Appendix 4). The original variable **local_heritage_sites** was intended as a measure of supply and was simply a count of heritage sites by local authority – it did not take account of the population in those authorities. In order to take account of local authority population size, producing a measure of supply density, the data were weighted by 2007 mid-year ONS population estimates (in thousands). This variable was highly skewed and so it was log transformed ($\log_e(X)$) in order to reduce skew.

Previous research on the drivers of engagement has suggested that multicolinearity among explanatory variables (e.g. **income** and **education**) leads to difficulties in isolating their effects on engagement (see section 3). In order to test whether multicolinearity was likely to

be a problem in our data, variance inflation factors (VIF) were calculated (see Appendix 6). None of the VIFs indicated that problematic multicolinearity was present in the explanatory variables (based on a rule of thumb that VIFs should not exceed 10).

Multivariate model selection was undertaken using robust single level logistic regression of the form[7]:

$$L_{ij} = Logit\left\{\Pr(y_{ij} = 1 \middle| x_{1j}, x_{2j}, x_{3ij}, x_{4ij}, ..., x_{nij})\right\}$$
$$= \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + ... + \beta_n x_{nij} \quad (2)$$

Where:

$i =$	Identifier for individual
$j =$	Identifier for area
$L_{ij} =$	Natural logarithm of odds of 'yes' response to a yes/no question
$y_{ij} =$	Response to yes/no question as a Boolean value
$\beta_0, \beta_1, \beta_2 ... =$	Coefficients of regression model
$x_{1j}, x_{2j} ... =$	Explanatory variables at area level
$x_{3ij}, x_{4ij}, ..., x_{nij} =$	Explanatory variables at individual level

The data were weighted using the variable **pweight1** (individual level weight, see Appendix 7 for more information on data). Early on, the effect of viewing heritage websites (**heritage_internet**) was identified as problematic. The effect size was much greater than for any other covariate. The likely reason for this very large effect was ambiguity in the causal direction, as looking at heritage websites might increase the likelihood of visiting a heritage site, but visiting a heritage site could also increase the likelihood of looking at heritage websites. Therefore, **heritage_internet** was removed from the model first.

Model selection then proceeded by removing terms in order of decreasing significance. A nominal significance level of 5% was used to determine whether terms were retained in the model. Once the terms remaining were all significant at the 5% level, interactions to identify additional variations in effect due to demographics, family structure, socio-economic group and free time were tested. Finally, the model was re-estimated with the final set of terms using the specification in equation (1) and weighted with **pweight1** and **pweight2** (area level weight). More details of the model selection process and Stata code used can be found in Appendix 7.

[7] The model specification varies from that used in the art model, shown in equation (1), as it was not possible to replicate the complexity of this multi-level model in the time available to undertake the heritage model. The model selection was undertaken using a single level robust logit model specification instead, with a random intercept being included in the final estimations. The risk to this two stage approach is that the model selection process may result in slightly different terms to a comparable process using multilevel models at every step of selection. However, this is most likely to affect the less influential effects and so, considering resource and time constraints, was found to give a good balance between robustness and feasibility. More details of this process can be found in Appendix 7.

3.4.2 Model results

Table 5 shows the final estimates from the multivariate modelling of visiting a **heritage** site. The bar chart at the side of the table demonstrates the strength of relationships – red bars for negative effects and green bars for positive effects. The bar sizes are proportional to either the odds ratio (where a variable is binary) or the odds ratio raised to the power of one standard deviation (where a variable is continuous or interval), demonstrating the relative effect of changing a binary variable from 0 to 1, or moving one standard deviation along the scale of a continuous variable.

The estimates reflected a similar picture to the bivariate analyses, with moderate to large effects being in evidence for variables measuring socio-economic indicators, media consumption, education, motor vehicle access and childhood experience. Additionally, however, there was a positive effect for local heritage site density, and there were interesting effects in evidence for age and being part of a BME group.

Table 5: Multivariate analysis of the factors associated with visits to heritage sites

Model terms	Odds ratio	Coefficient	Std. Err.	p-value	Less/More likely
Socio-economic characteristics					
Age	1.0316	0.0311	0.0101	0.002	
Age squared	0.9997	-0.0003	0.0001	0.003	
Member of a BME group vs. not	1.2285	0.2058	0.2788	0.460	
Member of a BME group interacted with age	0.9717	-0.0287	0.0074	0.000	
Religion: non-practicing vs. not religious	0.9080	-0.0965	0.1066	0.365	
Religion: practicing vs. not religious	1.2094	0.1902	0.1163	0.102	
High social-economic status vs. not	1.3877	0.3276	0.0712	0.000	
Income	1.0504	0.0492	0.0144	0.001	
Living as a couple vs. not	1.2737	0.2419	0.0915	0.008	
Children living in household vs. not	1.1859	0.1705	0.1385	0.218	
Children interaction with living as a couple	0.6474	-0.4348	0.1683	0.010	
Work status: part time vs. not working	0.9606	-0.0402	0.1293	0.756	
Work status: full time vs. not working	0.7956	-0.2287	0.1048	0.029	
Social housing tenant vs. not	0.7117	-0.3401	0.0906	0.000	
Educated to other higher level vs. degree level	0.6994	-0.3576	0.1466	0.015	
Educated to A-level vs. degree level	0.5266	-0.6413	0.1266	0.000	
Educated to trade apprentice level vs. degree level	0.4519	-0.7943	0.2168	0.000	
Educated to 5+ GCSEs A*-C level vs degree level	0.4403	-0.8203	0.1341	0.000	
Educated to <5 GCSEs A*-C level vs degree level	0.4177	-0.8730	0.1589	0.000	

Educated other qualifications vs degree level	0.3053	-1.1866	0.1833	0.000	
Educated to no qualifications vs. degree level	0.3182	-1.1451	0.1357	0.000	
Media access					
Watches history on TV vs. not	2.1858	0.7820	0.0786	0.000	
Has access to internet vs. not	1.5525	0.4399	0.0866	0.000	
Hours of TV watched per day	0.8976	-0.1080	0.0293	0.000	
Accessibility of engagement					
Has access to a motor vehicle vs. not	1.3080	0.2685	0.0913	0.003	
Log of population weighted local heritage sites	1.1653	0.1530	0.0342	0.000	
Childhood experience					
Average no. times taken to heritage sites p.a. as child	1.1598	0.1482	0.0193	0.000	
constant	-	*0.2045*	*0.2831*	*0.470*	

number of individual respondents = 10063

number of local authorities = 346

variance of random intercept = 0.2014

standard error of random intercept = 0.0422

(Note: the bar chart is approximate - the sizes of bars are proportional to the odds ratios raised to the power of one for dummy variables or the odds ratios raised to the power of one standard deviation for continuous variables)

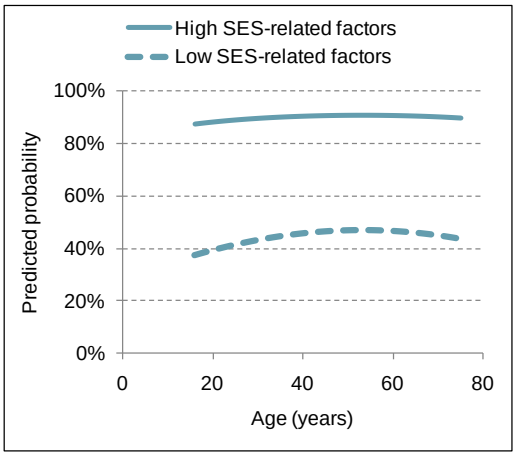
Predictions were created using the model for various combinations of the explanatory variables. *Figure 11a* shows a comparison of the predicted probability of visiting heritage sites by the age of individuals who are high on all factors related to socio-economic advantage with those who are low on the same factors. The dashed line represents individuals who are of low socio-economic status, have an annual income of £10,000 to £14,999, who watch four hours of television per night, are social housing tenants, do not have access to a motor vehicle and have an education level of fewer than five A*-C grades at GCSE. The solid line represents individuals who are of high socio-economic status, have an annual income of £50,000 or more, who watch two hours of television per night, are not social housing tenants, have access to a motor vehicle and have an education level of degree or above. Combined together, these factors have a very large impact on the predicted probability of heritage engagement, with 90% of high socio-economic individuals predicted to engage compared with 40%-50% of low socio-economic individuals.

Media consumption was also shown to be strongly associated with heritage engagement. *Figure 11b* shows a comparison of the predicted probability of visiting heritage sites by age for individuals who are media rich with those who are media poor. The solid line represents individuals who have access to the internet, watch history TV programmes and watch two hours of television per day. The dashed line represents individuals who watch no television and do not have access to the internet. Combined together, these factors have a moderate

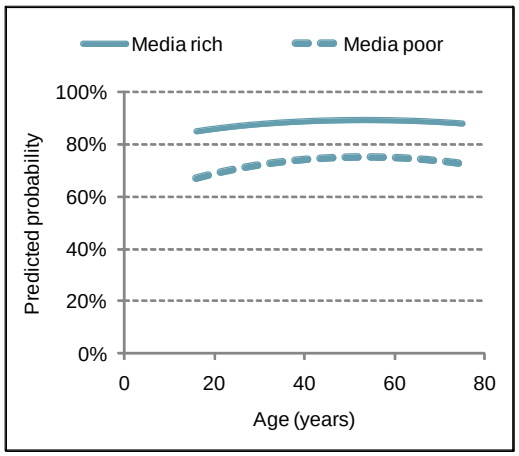
impact on the predicted probability of engaging with art, with around 90% of media-rich individuals predicted to engage compared to 70% of media-poor individuals.

Figure 11: Predicted probability of visiting heritage sites by socio-demographic factors

11a. High and low SES-related factors



11b. Media consumption



11c. BME and non-BME groups

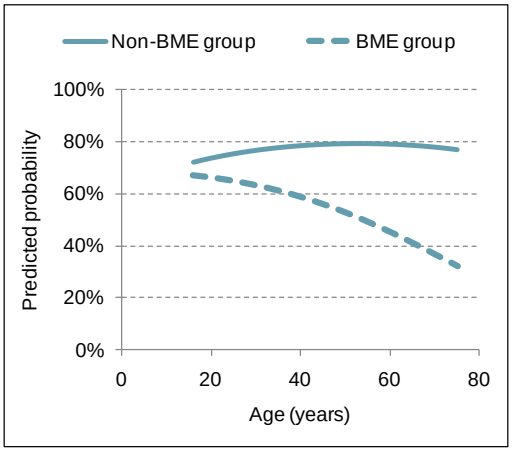


Figure 11c shows a comparison of the predicted probability of visiting heritage sites by age for individuals in a BME group with those who are not. The dashed line represents individuals who are in a BME group, the solid line represents individuals who are not. Interestingly, younger BME individuals are predicted to be almost as likely to engage with heritage as non-BME individuals, but this predicted engagement declines steeply with age. This observation suggests interesting inter-generation effects in the BME population. For instance, perhaps the tastes of younger (perhaps third generation) members of the BME group are converging with those of the non-BME population?

There were also effects that reflected the geographic distance of heritage sites. Predicted engagement increased for individuals with access to a motor vehicle and was also greater in local authorities where there was a higher density of heritage sites. Also, amount of free time available was found to be important, with engagement reducing as working status increased from part-time to full-time.

3.4.3 Model testing and diagnostics

Software for multilevel modelling is relatively new and is developing as computers become more powerful and are able to cope with the large number of calculations needed to estimate these types of model. As a result, the number of fit and diagnostic tests available for use with weighted multilevel models is limited. In order to test the final heritage model thoroughly, a single-level logistic regression model with robust standard errors was constructed to mimic the multilevel model. Apart from not containing a random intercept parameter, this model contained the same terms as the multilevel heritage model. In the model, the random intercept term was significant at the 5% level, thus a single-level model does not fit the data as well as the multilevel model. Therefore, if the single-level model was found to fit the data well, using the tests available for single-level unweighted models, it is reasonable to assume that the multilevel model would also fit the data well.

A Hosmer-Lemeshow test of the single-level model suggested that it was a good fit to the data (50 groups, $Chi^2(48)=54.99$, $p=0.2270$). However, a model specification link test suggested the model could possibly be improved ($p=0.042$) by addition of a square term. This result might be due to the reduced flexibility of the single-level model as compared to the multilevel model, but it is also possible that alternative models could be built that would significantly improve fit. Diagnostic plots were produced using the multilevel heritage model. A plot of standardised residuals at the local authority level is shown in *Figure 12* and a plot of deviance residuals for the respondent level is shown in *Figure 13*. The plots do not demonstrate any abnormalities that would cause concern. Further information on model testing and diagnostics can be found in Appendix 7.

Figure 12: Local authority level standardised residuals

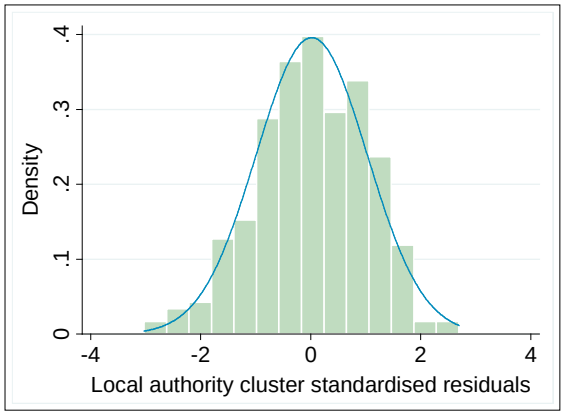
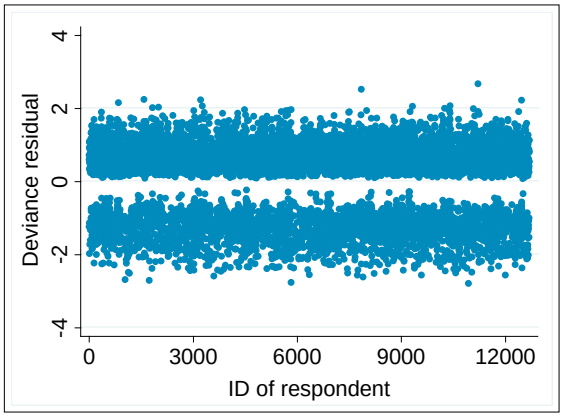


Figure 13: Respondent level deviance residuals



Estimates of explanatory power for logistic regression (pseudo-R^2 measures) do not have the same meaning as those for regression models where the response variable is continuous and are not generally regarded as robust by statisticians. However, two estimates were calculated using the single level robust weighted model to give an indication of explanatory power:

- McFadden's R^2 – this resulted in an estimate of a 17.3% reduction in error.
- McKelvey and Zavoina's R^2 – this resulted in an estimate of 30.8% reduction in error.

3.5 Participating in sport

3.5.1 Model development

The sport model was built using the response variable **sport** which was encoded from the survey data and based on the definition of Sport England's "1 million sport" indicator: three times 30 minutes per week of moderate intensity sport in the past month, excluding any walking and low-frequency (less than once a week) cycling. The model was constructed by populating the conceptual framework (Appendix 1). The variables used to construct this first model are described in *Table 6*. More information on the data is available in Appendices 3 and 4.

Table 6: Dataset used to build sport model

Variable name	Description
access	Accessibility to services index from National Indicator set (percent) – percentage of people who can reach selected core services and facilities within 15 minutes by public transport and/or walking (National Indicator 175)
age	Age in years
alcohol	Days alcohol drunk per week (categorical from 0=none to 4=every day)
BMEgroup	Member of ethnic minority group (1=yes, 0=no)
child_encouraged_sport	How much encouraged to do sport when growing up (categorical from 0=not to 2=a lot)
children	Children living in household (1=yes, 0=no)
community_cohesion	% of people who feel that they belong to their neighbourhood
coupled	Living as part of a couple (1=yes, 0=no)
cycles	Cycles to get from place to place (1=yes, 0=no)
education	Highest educational qualification held (categorical from 1=degree to 8=no qualifications)
highSES	Member of high socio-economic group NS-SEC 1-4 (1=yes, 0=no)
income	Highest income in household (interval from 0=£0 to 12=£50,000 or more)
internet	Has access to internet (1=yes, 0=no)
limiting_illness	Has illness or disability which limits activities (1=yes, 0=no)
live_sport_tv	Watches live sport coverage television programmes (1=yes, 0=no)
local_parks_satisfaction	% of people who are very or fairly satisfied with parks and open spaces
local_sports_facilities	Population weighted count of sport assets in area – including the following categories: Grass Pitch, Sports Hall, Swimming Pool, Golf, Synthetic Turf Pitch, Health and Fitness Suite, Athletics Tracks, Indoor Tennis Centre, Indoor Bowls, Ice Rinks, Ski Slopes
local_sports_satisfaction	% of people who are very or fairly satisfied with sport/leisure facilities
motor_vehicle	Has access to a motor vehicle (1=yes, 0=no)
newspaper	Reads daily newspaper at least three times per week (1=yes, 0=no)
other_sport_tv	Watches other sport television programmes (1=yes, 0=no)
radio	Radio available in household (1=yes, 0=no)
religious	Religiosity (0=not religious, 1=non-practising, 2=practising)
sex	Sex of respondent (1=male, 0=female)
social_housing	Is a social housing tenant (1=yes, 0=no)

Variable name	Description
sport *(response variable)*	3x30 minutes moderate intensity sport per week in last month ('1 million sport' indicator)
sport_facility_nearby	Can get to a sports facility within 20 minutes (1=yes, 0=no)
sport_internet	Visited sport website in last 12 months (1=yes, 0=no)
sports_influence	Has influence over local sporting facilities (categorical from 0=no to 2=a lot)
tv_hours	Hours of television watched per day on average
work_status	Work status (0=not working, 1=part-time, 2=full-time)

Bivariate analyses of the explanatory variables with doing **sport** had identified many significant associations (see Appendix 5). Some of the strongest associations involved age and sex, childhood experience, limiting illness, TV viewing and internet access, cycling for travel, socio-economic factors, proximity of and influence over sports facilities. *Figure 14* shows a plot of doing **sport** against age. There is a very steep trend of a reducing probability of doing sport with age.

Figure 14: Proportion doing sport by age

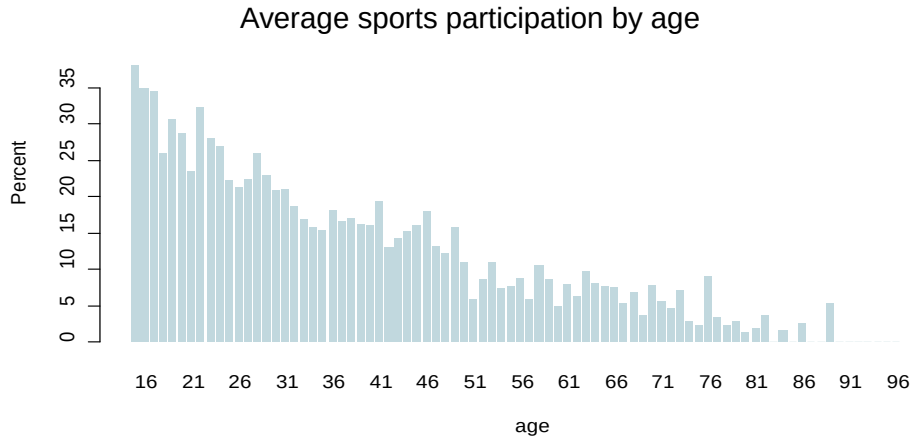
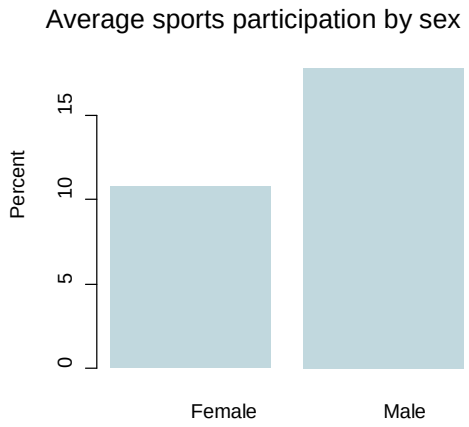


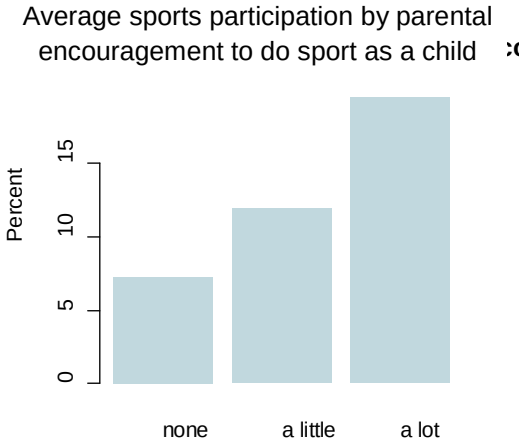
Figure 15 shows the probability of doing **sport** by sex. It is clear that there is a difference in average levels of sport engagement between the sexes, with males being a third more likely to do sport defined by the “1 million” indicator.

Figure 15: Proportion doing sport by sex



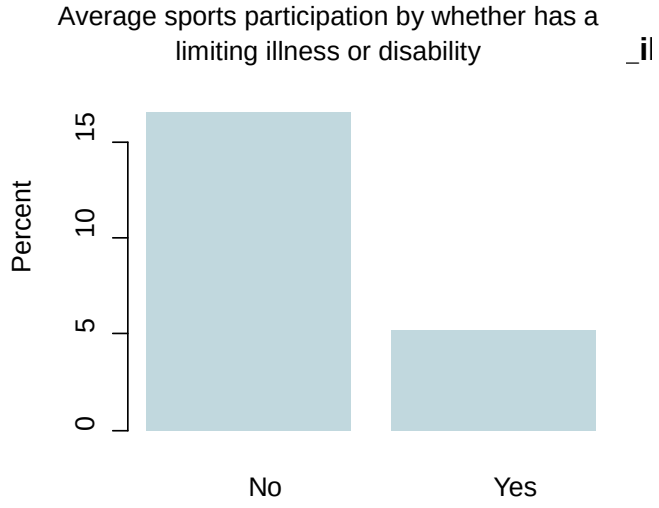
Childhood encouragement seemed to have a strong positive effect on sport engagement. *Figure 16* shows the probability of doing **sport** by level of childhood encouragement.

Figure 16: Proportion doing sport by level of childhood encouragement in sport



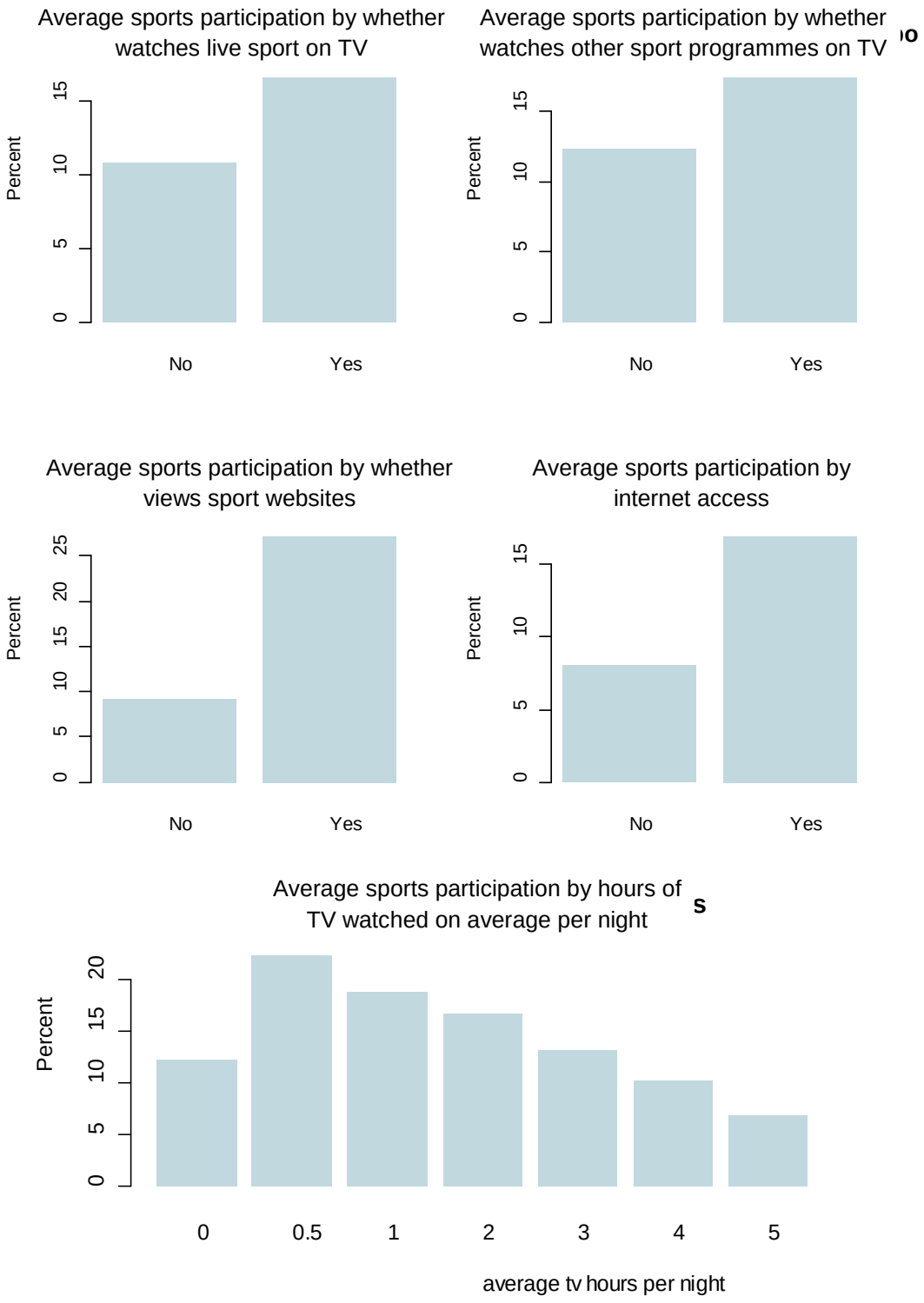
As expected, having a limiting illness negatively impacted on the average probability of doing sport. *Figure 17* shows that the average probability of doing sport for those individuals with a limiting illness was less than one third the level of those without a limiting illness.

Figure 17: Proportion doing sport by limiting illness



Measures of media consumption exhibited strong positive associations with doing sport, except for the number of hours of television watched on average per night, which was negatively associated with doing sport. Furthermore, this data points to a potential associate between participating in sport and viewing sport. *Figure 18* shows plots of these associations.

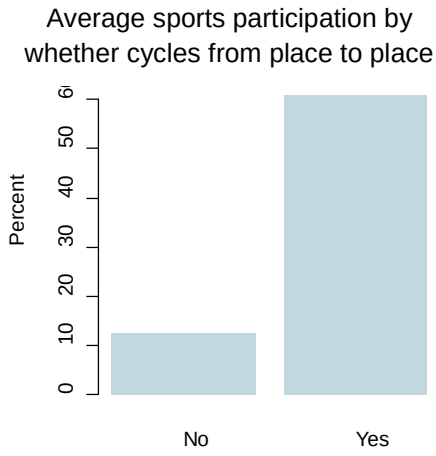
Figure 18: Proportion doing sport by internet and television consumption



An interesting relationship between cycling for travel purposes and meeting the "1 million sport" level of engagement was apparent. This suggests that a majority of people who cycle for travel either meet the three times 30 minutes per week moderate intensity level through this cycling or also do other sporting activities in order to meet this level. *Figure 19* shows the very large difference in average probability of doing sport by whether the individual

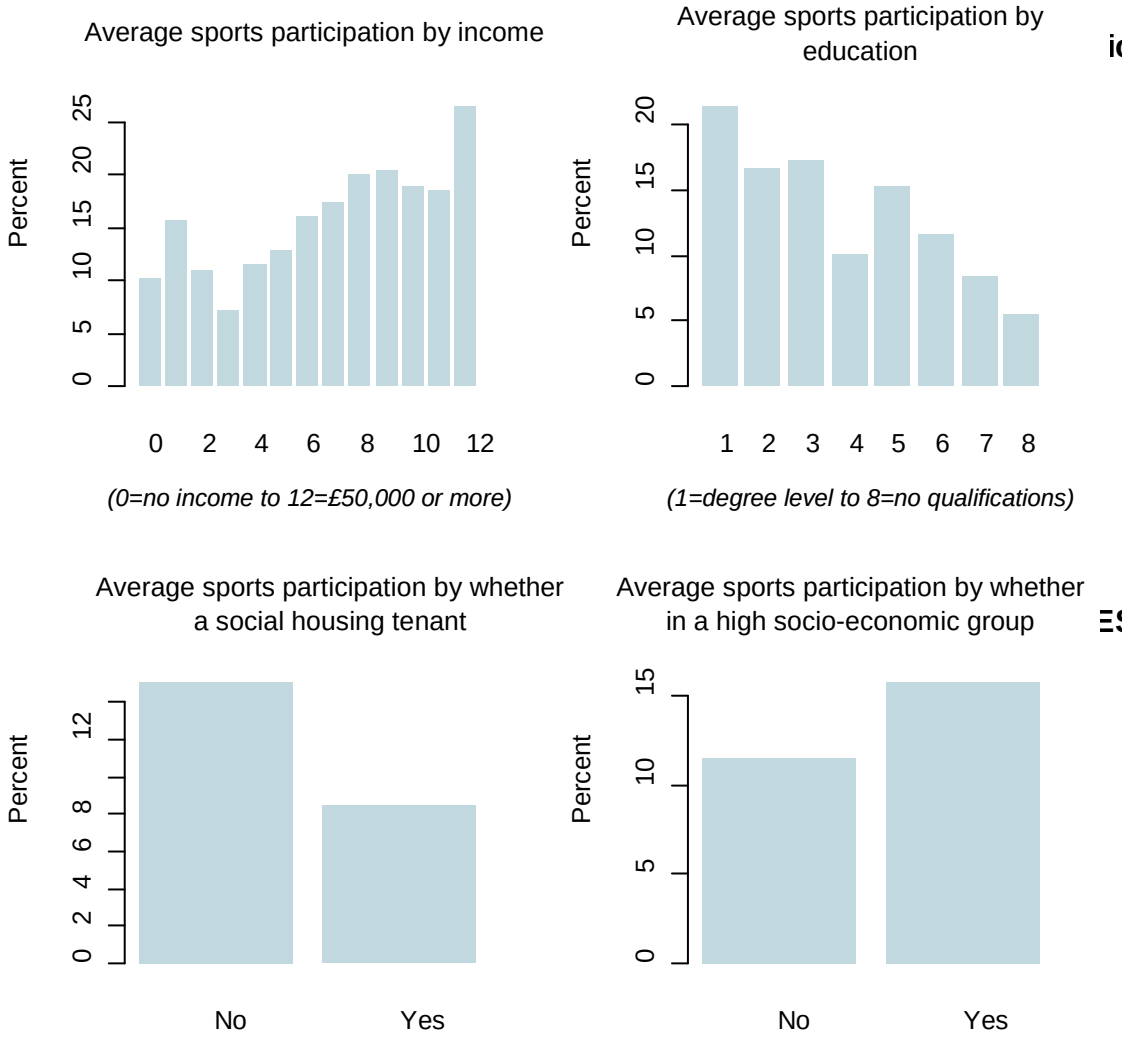
cycles for travel or not. Due to the possibility of confounding of **cycles** as an explanatory variable with the response variable **sport**, it was decided at this point to exclude this variable from subsequent regression modelling.

Figure 19: Proportion doing sport by whether a persons cycles or not



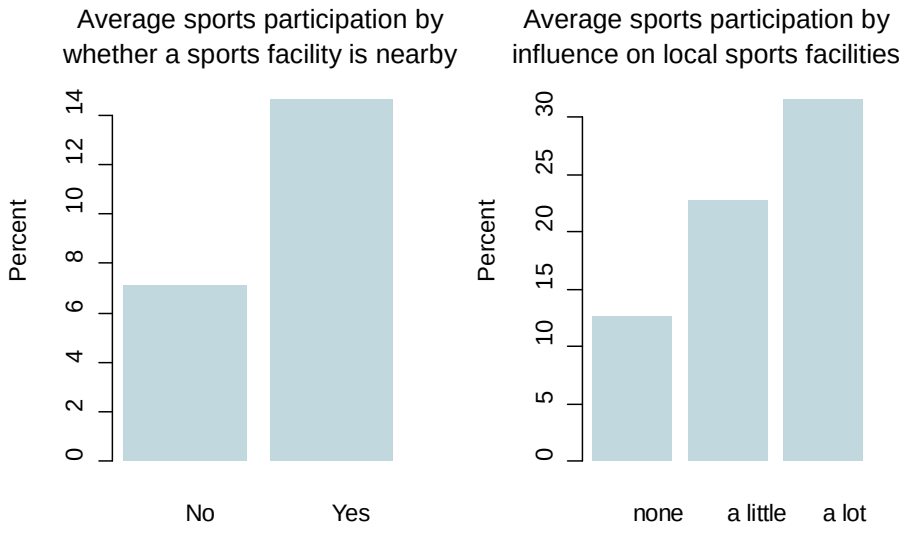
Measures related to socio-economic status and education exhibited positive and negative bivariate associations with the probability of doing sport. Doing **sport** was positively associated with income and socio-economic status, and negatively associated with lower education levels and social housing. *Figure 20* shows plots of the probability of doing sport by income, education, social housing and socio-economic status.

Figure 20: Proportion doing sport by income, education level, social housing and socio-economic status



There also seemed to be an association between both reported sports facility proximity and the degree of individual influence on sporting facilities and the probability of doing sport. *Figure 21* shows sport engagement by **sport_facility_nearby** and **sports_influence**. These observations could suggest that increasing the proximity of sporting facilities improves engagement rates. However, it is important to note that this is based on a bivariate association – a range of other factors will influence engagement that aren't included in the analysis – and the line of causation may run from interest in sport to locating near sporting facilities, rather than the other way round.

Figure 21: Proportion doing sport by sport facility proximity and individual influence on facilities



Interval and ratio variables were plotted to determine whether any exhibited a high degree of skew, which could lead to instability in model estimation (see Appendix 4). The original variable **local_sports_facilities** was intended as a measure of supply and was simply a count of sport facilities by local authority – it did not take account of the population in those authorities. In order to take account of local authority population size, producing a measure of supply density, the data were weighted by 2007 mid-year ONS population estimates (in thousands).

Previous research on the drivers of engagement has suggested that multicolinearity among explanatory variables (e.g. **income** and **education**) leads to difficulties in isolating their effects on engagement (see section 3). In order to test whether multicolinearity was likely to be a problem in our data, variance inflation factors (VIF) were calculated (see Appendix 6). None of the VIFs indicated that problematic multicolinearity was present in the explanatory variables (based on a rule of thumb that VIFs should not exceed 10).

The model constructed was of the form:

$$L_{ij} = Logit\left\{\Pr(y_{ij} = 1 \middle| x_{1j}, x_{2j}, x_{3ij}, x_{4ij}, ..., x_{nij}, \zeta_j)\right\}$$
$$= \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + ... + \beta_n x_{nij} + \zeta_j \quad (3)$$

Where:

$i =$	Identifier for individual
$j =$	Identifier for area
$L_{ij} =$	Natural logarithm of odds of 'yes' response to a yes/no question
$y_{ij} =$	Response to yes/no question as a Boolean value
$\beta_0, \beta_1, \beta_2 ... =$	Coefficients of regression model
$x_{1j}, x_{2j} ... =$	Explanatory variables at area level
$x_{3ij}, x_{4ij}, ..., x_{nij} =$	Explanatory variables at individual level
$\zeta_j =$	Random intercept for area

The majority of the data used to construct the models were at the individual level. Due to the difficulty in including all relevant variables at the area level, the model specification included a random intercept term. This term increases the flexibility of the logistic model to give a separate intercept for each Local Authority, thereby allowing for heterogeneity in engagement levels due to unmeasured area level factors and increasing the accuracy of estimates.

The data were weighted using the variables **pweight1** (individual level weight) and **pweight2** (area level weight, see Appendix 7 for more information on data). Model selection proceeded by removing terms in order of decreasing significance. A nominal significance level of 5% was used to determine whether terms were retained in the model. Once the terms remaining were all significant at the 5% level, interactions to identify additional variations in effect due to demographics, family structure, socio-economic group and free time were tested. More details of the model selection process and Stata code used can be found in Appendix 7.

3.5.2 Model results

Table 7 shows the final estimates from the multivariate modelling of the factors associated with doing **sport**. The bar chart at the side of the table demonstrates the strength of relationships – red bars for negative effects and green bars for positive effects. The bar sizes are proportional to either the odds ratio (where a variable is binary) or the odds ratio raised to the power of one standard deviation (where a variable is continuous or interval), demonstrating the relative effect of changing a binary variable from 0 to 1, or moving one standard deviation along the scale of a continuous variable

Table 7: Multivariate model of the factors associated with doing sport

Model terms	Odds ratio	Coefficient	Std. Err.	p-value	Less/more likely
Socio-economic characteristics					
Sex: male vs. female	3.1200	1.1378	0.2425	0.000	
Sex interaction with age	0.9800	-0.0202	0.0055	0.000	
Age	0.9731	-0.0272	0.0040	0.000	
Income	1.0369	0.0362	0.0157	0.021	
Work status: part time vs. not working	0.8283	-0.1884	0.1517	0.214	
Work status: full time vs. not working	0.7531	-0.2836	0.1268	0.025	
Educated to other higher level vs. degree	0.8509	-0.1615	0.1454	0.267	
Educated to A-level vs. degree	0.6907	-0.3701	0.1362	0.007	
Educated to trade apprentice level vs. degree	0.6664	-0.4058	0.2308	0.079	
Educated to 5+ GCSEs A*-C level vs degree	0.7822	-0.2457	0.1441	0.088	
Educated to <5 GCSEs A*-C level vs degree	0.5499	-0.5980	0.2196	0.006	
Educated other qualifications vs degree	1.1451	0.1355	0.2751	0.622	
Educated to no qualifications vs. degree	0.6247	-0.4705	0.1913	0.014	
Member of a BME group vs. not	0.6630	-0.4110	0.1439	0.004	
Children living in household vs. not	0.6720	-0.3975	0.0863	0.000	
Illness or disability vs. not	0.4961	-0.7011	0.1470	0.000	
Media access					
Visited sport website in last year vs. not	1.8048	0.5905	0.1179	0.000	
Watches live sport on TV vs. not	1.2725	0.2410	0.0923	0.009	
Hours of TV watched per day	0.8955	-0.1104	0.0330	0.001	
Childhood experience					
Encouraged to play sport as child: a lot vs. not	1.5987	0.4692	0.1252	0.000	
Encouraged to play sport as child: some vs. not	1.2262	0.2040	0.1411	0.148	
Influence over provision					
Influence over sport facilities: a lot vs. none	2.7571	1.0142	0.2642	0.000	
Influence over sport facilities: some vs. none	1.2986	0.2613	0.1161	0.024	
constant	-	*-0.7985*	*0.2659*	*0.003*	

number of individual respondents = 9996
number of local authorities = 346
variance of random intercept = 0.2509
standard error of random intercept = 0.0599

(Note: the bar chart is approximate - the sizes of bars are proportional to the odds ratios raised to the power of one for dummy variables or the odds ratios raised to the power of one standard deviation for continuous variables)

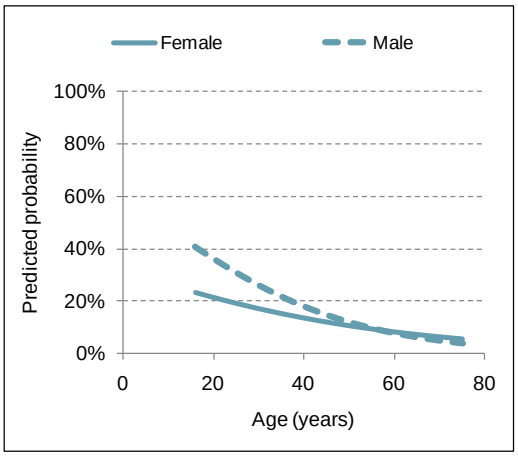
The estimates demonstrated the very large impact of sex and influence on sports facilities on the probability of doing sport. Other large effects included childhood encouragement, visiting sport websites, watching live sport on TV and having a limiting illness. Having children and being part of a BME group were moderately negatively associated with “1 million sport” engagement.

Predictions were created using the model for various combinations of the explanatory variables. *Figure 22a* shows predicted probability of doing sport by age for males as compared with females. Predicted probability of doing sport for females is approximately half that of males for individuals of around 20 years of age, but this difference reduces with the age of the respondents, with the difference disappearing at around 55 years of age.

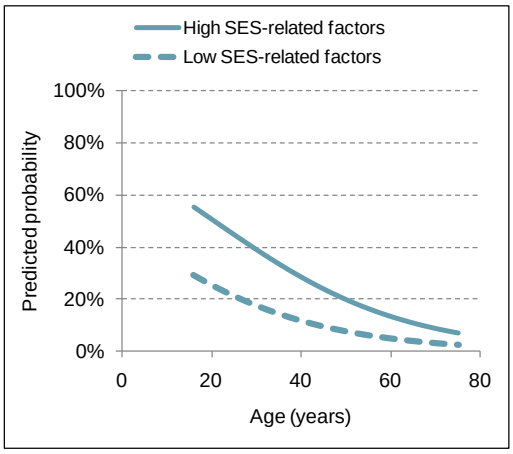
Figure 22b shows a comparison of the predicted probability of doing sport by age for individuals who are high on all factors related to socio-economic advantage with those who are low on the same factors. The dashed line represents individuals who have an annual income of £10,000 to £14,999, who watch four hours of television per night and have an education level of fewer than five A*-C grades at GCSE. The solid line represents individuals who have an annual income of £50,000 or more, who watch two hours of television per night and have an education level of degree or above. Combined together, these factors cause the predicted probability of sport engagement for low socio-economic individuals to be approximately half that of high socio-economic individuals, across the age range.

Figure 22: Predicted probability of doing sport by socio-demographic factors

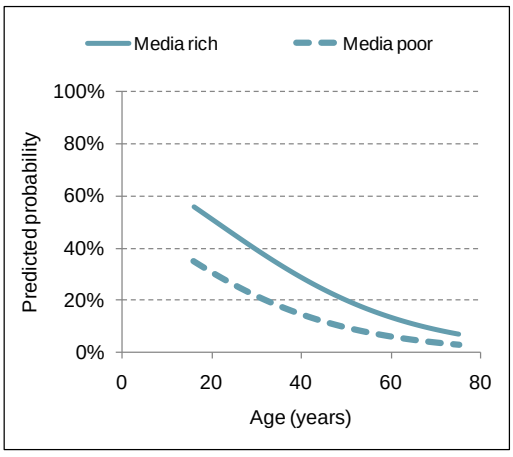
22a. Male and female



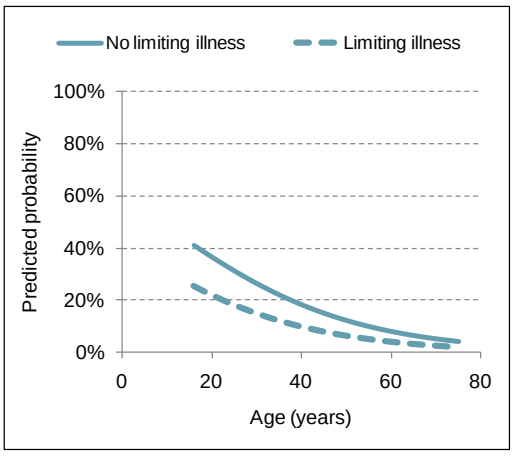
22b. High and low SES



22c. Media consumption



22d. Limiting illness



Media consumption was shown to have a strong association with the probability of doing sport. *Figure 22c* above shows a comparison of the predicted probability of doing sport by age for individuals who are sport-media-rich with those who are sport-media-poor. The solid line represents individuals who watch live sport TV programmes and visit sport websites. The dashed line represents individuals who do not watch live sport TV programmes and do not visit sport websites. Combined together, these factors have a substantial impact on the predicted probability of sport engagement, though not quite as large as the effect of combined socio-economically related factors. The analysis suggests that greater exposure to media (perhaps as a result of greater watching of sport) may increase the probability that young people participating in sport by 1.5 times.

Figure 22d above shows a comparison of the predicted probability of doing sport by age for individuals who have a limiting illness with those who do not. The dashed line represents individuals who have a limiting illness, the solid line represents individuals who do not. The effect, although large, is not as big as might be expected, with approximately two-thirds as many people with a limiting illness engaging in sport as those without an illness. As people age, the effect of limiting illness on participation rates declines, as most older people don't meet the threshold for sports participation.

3.5.3 Model testing and diagnostics

Software for multilevel modelling is relatively new and is developing as computers become more powerful and are able to cope with the large number of calculations needed to estimate these types of model. As a result, the number of fit and diagnostic tests available for use with weighted multilevel models is limited. In order to test the final sport model thoroughly, a single-level logistic regression model with robust standard errors was constructed to mimic the multilevel model. Apart from not containing a random intercept parameter, this model contained the same terms as the multilevel sport model. In the model, the random intercept term was significant at the 5% level, thus a single-level model does not fit the data as well as the multilevel model. Therefore, if the single-level model was found to fit the data well, using the tests available for single-level unweighted models, it is reasonable to assume that the multilevel model would also fit the data well.

A Hosmer-Lemeshow test of the single-level model suggested that it was a good fit to the data (50 groups, $Chi^2(48)=45.72$, $p=0.5667$) and this was supported by a model specification link test ($p=0.133$). Diagnostic plots were produced using the multilevel sport model. A plot of standardised residuals at the local authority level is shown in *Figure 23* and a plot of deviance residuals for the respondent level is shown in *Figure 24*. The plots do not demonstrate any abnormalities that would cause concern. Further information on model testing and diagnostics can be found in Appendix 7.

Figure 23: Local authority level standardised residuals

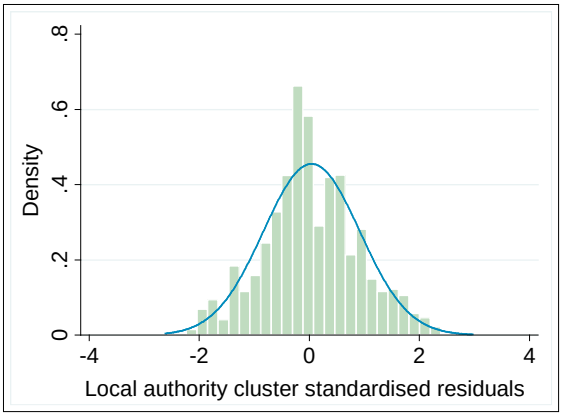
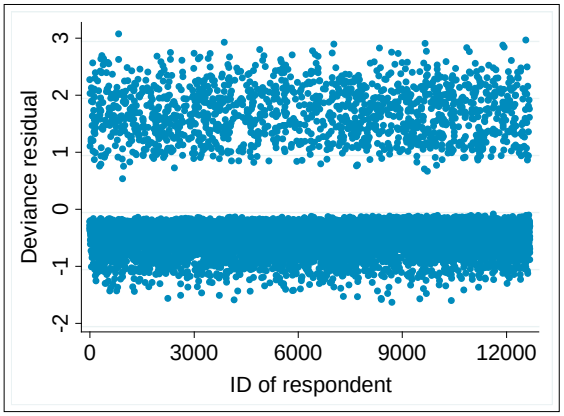


Figure 24: Respondent level deviance residuals



Estimates of explanatory power for logistic regression (known as pseudo-R^2 measures) do not benefit from the straightforward interpretation found in continuous regression models because the response variables can only take the values 0 or 1. As a result, although many different versions are available, statisticians do not view any of these measures as particularly robust and they only serve as weak analogues of the more familiar R^2 in continuous regression. Generally, pseudo-R2 measures return low values of explanatory power of between 10% and 30%. Despite these reservations, two estimates were calculated using the single level robust weighted model to give comparable indications of explanatory power:

- McFadden's R^2 – this resulted in an estimate of a 14.7% reduction in error using the model to predict engagement as compared to the overall mean.
- McKelvey and Zavoina's R^2 – this resulted in an estimate of 27.2% reduction in error using the model to predict engagement as compared to the overall mean.

3.6 Visiting a library

3.6.1 Model development

An initial analysis was undertaken on the factors associated with visiting a museum, library or an archive (MLA). The initial results from this model demonstrated the likely problems arising from modelling a response variable that encompassed engagement with museums, libraries and archives all together. It was therefore decided to separate the component parts of MLA. As the proportion of Taking Part respondents visiting archives was low (approximately 5%), models were run for visiting museums and for visiting libraries.

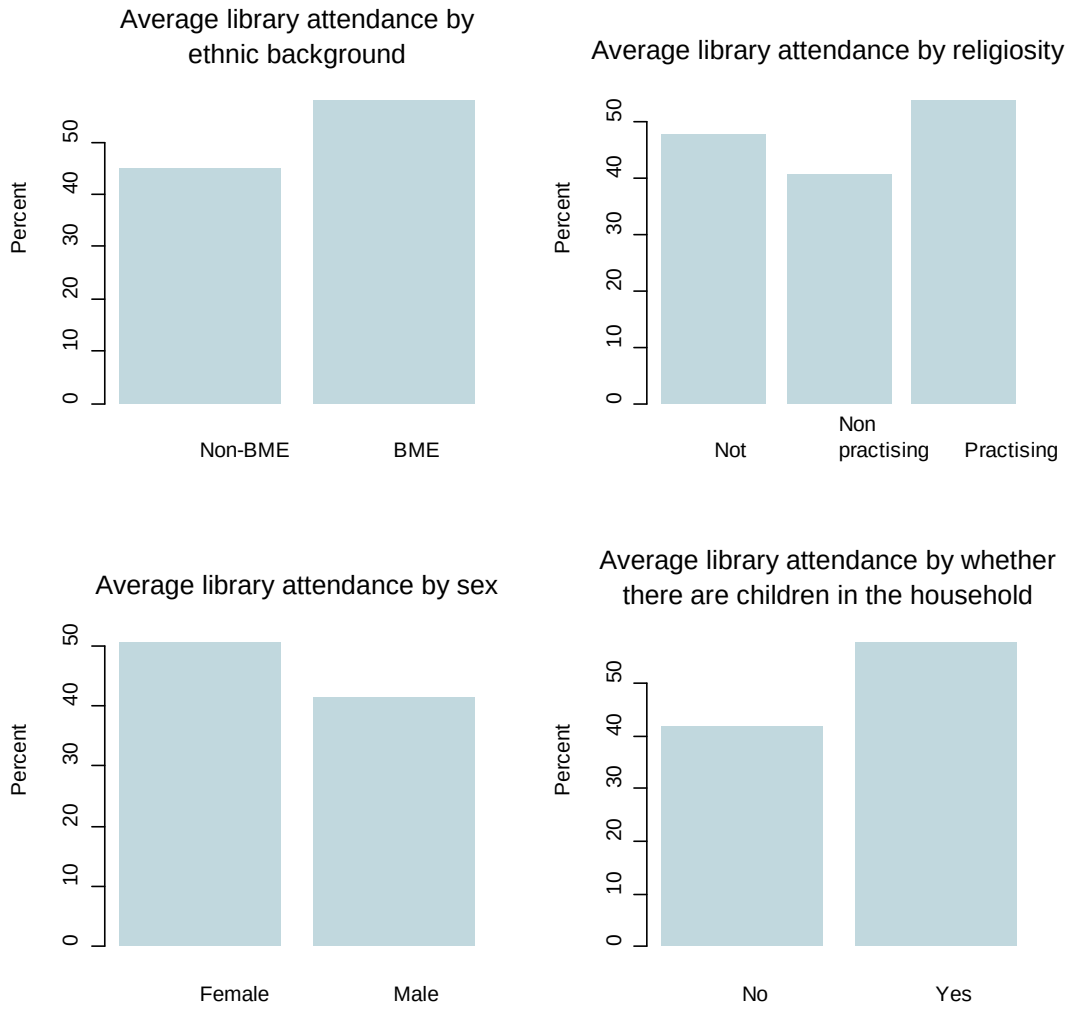
The library model was built using the response variable **library** – whether the respondent had visited a library in the past 12 months. The model was constructed by populating the conceptual framework (Appendix 1). The variables used to construct this model are described in *Table 8*. More information on the data is available in Appendices 3 and 4

Table 8: Variables used to build library model

Variable name	Description
access	Accessibility to services index from National Indicator set (percent) – percentage of people who can reach selected core services and facilities within 15 minutes by public transport and/or walking (National Indicator 175)
age	Age in years
BMEgroup	Member of ethnic minority group (1=yes, 0=no)
child_encouraged_read	How much encouraged to do read when growing up (categorical from 0=not to 2=a lot)
child_library_visit	Average times per year taken to libraries when growing up
children	Children living in household (1=yes, 0=no)
community_cohesion	% of people who feel that they belong to their neighbourhood
coupled	Living as part of a couple (1=yes, 0=no)
cultural_influence	Has influence over local cultural facilities (categorical from 0=no to 2=a lot)
cycles	Cycles to get from place to place (1=yes, 0=no)
education	Highest educational qualification held (categorical from 1=degree to 8=no qualifications)
highSES	Member of high socio-economic group NS-SEC 1-4 (1=yes, 0=no)
income	Highest income in household (interval from 0=£0 to 12=£50,000 or more)
internet	Has access to internet (1=yes, 0=no)
library	Visited library in last 12 months (1=yes, 0=no)
library_internet	Visited library website in last 12 months (1=yes, 0=no)
limiting_illness	Has illness or disability which limits activities (1=yes, 0=no)
local_libraries	Population weighted count of libraries in area – categorised by SIC(92) classification code 9251 (Library And Archive Activities)
local_library_satisfaction	% of people who are very or fairly satisfied with libraries
motor_vehicle	Has access to a motor vehicle (1=yes, 0=no)
newspaper	Reads daily newspaper at least three times per week (1=yes, 0=no)
radio	Radio available in household (1=yes, 0=no)
religious	Religiosity (0=not religious, 1=non-practising, 2=practising)
sex	Sex of respondent (1=male, 0=female)
social_housing	Is a social housing tenant (1=yes, 0=no)
tv_hours	Hours of television watched per day on average
work_status	Work status (0=not working, 1=part-time, 2=full-time)

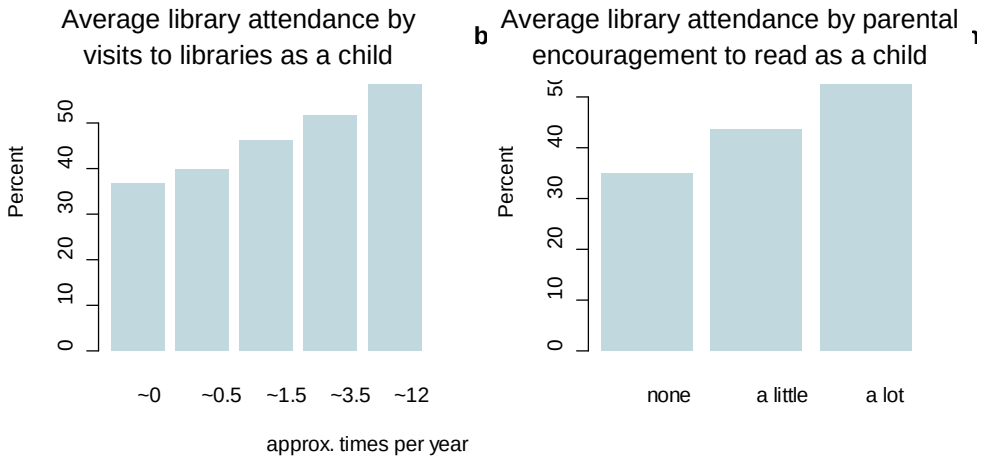
Bivariate analyses of the explanatory variables with visiting a **library** identified many significant associations (see Appendix 5). Some of the strongest associations involved socio-demographic characteristics, children living in the household, childhood experience, education and media factors. *Figure 25* shows plots of library engagement by the socio-demographic characteristics ethnicity, sex and religiosity, and also whether there are children living in the household.

Figure 25: Proportion of population visiting a library by ethnicity, religiosity, sex and children in household



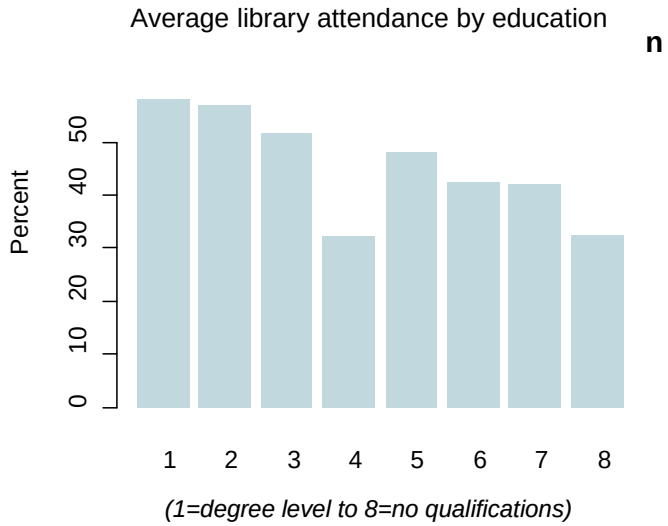
Measures related to childhood experience exhibited positive bivariate associations with visiting a library. Visiting a library was positively associated both with reported frequency of visiting libraries as a child and with the degree to which the respondent recalls being encouraged to read as a child. These associations are shown in *Figure 26*.

Figure 26: Proportion visiting a library by childhood experience of visiting libraries and being encouraged to read by parents



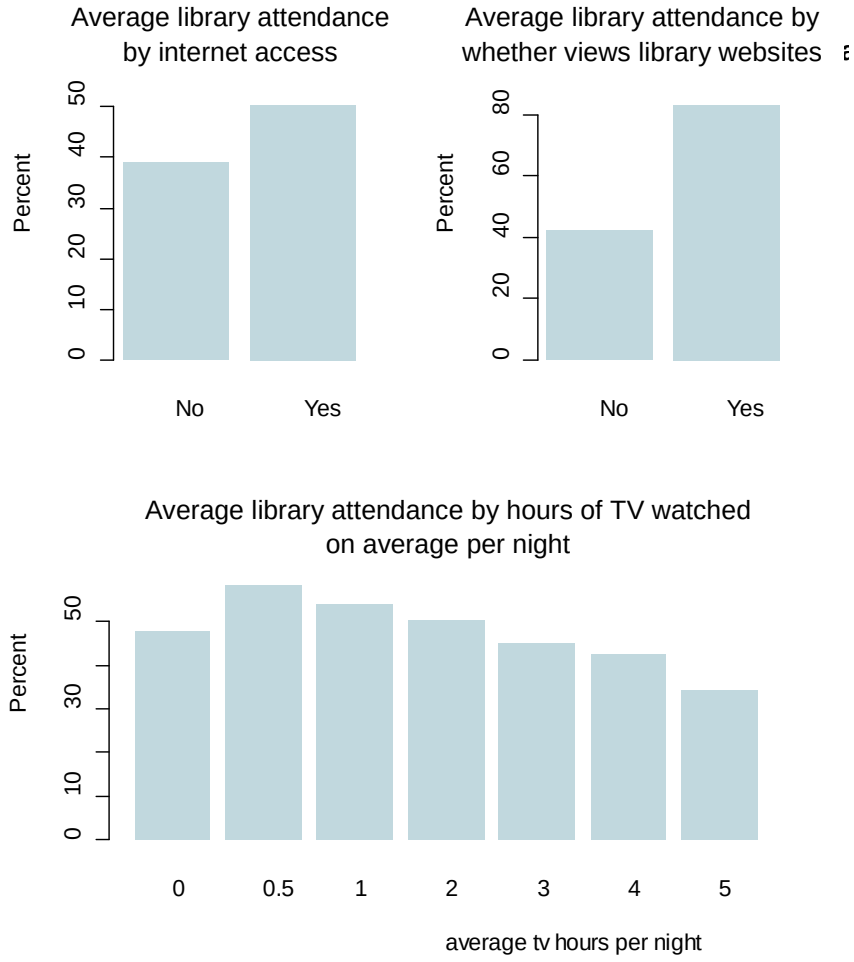
Level of education also exhibited a positive bivariate association with visiting a library. *Figure 27 shows* a plot of proportion of people visiting a library by education level. Lower numbered categories indicate a higher level of education, with category 1 representing degree level and category 8 representing no qualifications.

Figure 27: Proportion visiting a library by level of education



Measures of internet consumption exhibited strong positive associations with library engagement. Having access to the internet and visiting library websites were both associated with increased average levels of engagement. However, engagement was negatively associated with hours of TV viewing. *Figure 28* shows plots of library engagement against internet access, visiting library websites and average TV viewing hours per week.

Figure 28: Proportion visiting a library by internet access, visiting a library website and TV viewing hours per week



Interval and ratio variables were plotted to determine whether any exhibited a high degree of skew, which could lead to instability in model estimation (see Appendix 4). The original variable **local_libraries** was intended as a measure of supply and was simply a count of libraries by local authority – it did not take account of the population in those authorities. In order to take account of local authority population size, producing a measure of supply density, the data were weighted by 2007 mid-year ONS population estimates (in thousands). This variable was highly skewed and so it was log transformed ($\log_e(X+0.01)$) in order to reduce skew.

Previous research on the drivers of engagement has suggested that multicolinearity among explanatory variables (e.g. **income** and **education**) leads to difficulties in isolating their effects on engagement (see section 3). In order to test whether multicolinearity was likely to be a problem in our data, variance inflation factors (VIF) were calculated (see Appendix 6). None of the VIFs indicated that problematic multicolinearity was present in the explanatory variables (based on a rule of thumb that VIFs should not exceed 10).

Multivariate model selection was undertaken using robust single level logistic regression of the form[8]:

$$\begin{aligned} L_{ij} &= Logit\left\{\Pr(y_{ij}=1 \middle| x_{1j}, x_{2j}, x_{3ij}, x_{4ij}, ..., x_{nij})\right\} \\ &= \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + ... + \beta_n x_{nij} \end{aligned} \quad (4)$$

Where:

$i =$	Identifier for individual
$j =$	Identifier for area
$L_{ij} =$	Natural logarithm of odds of 'yes' response to a yes/no question
$y_{ij} =$	Response to yes/no question as a Boolean value
$\beta_0, \beta_1, \beta_2 ... =$	Coefficients of regression model
$x_{1j}, x_{2j} ... =$	Explanatory variables at area level
$x_{3ij}, x_{4ij}, ..., x_{nij} =$	Explanatory variables at individual level

The data were weighted using the variable **pweight1** (individual level weight, see Appendix 7 for more information on data). Model selection proceeded by removing terms in order of decreasing significance. A nominal significance level of 5% was used to determine whether terms were retained in the model. Once the terms remaining were all significant at the 5% level, interactions to identify additional variations in effect due to demographics, family structure, socio-economic group and free time were added. Finally, the model was re-estimated with the final set of terms using the specification in equations (1) & (3) and weighted with **pweight1** and **pweight2** (area level weight).

During the model selection process, the effect of viewing library websites (**library_internet**) was identified as problematic. The effect size was much greater than for any other covariate. The likely reason for this very large effect was ambiguity in the causal direction, as visiting library websites might be a result of visiting a library, rather than be a precursor to library visiting. Therefore, **library_internet** was removed from the model, and some previously

[8] Model selection was undertaken using robust single level logistic regression rather than the multilevel specification shown in equations (1) & (3). The risk to this two stage approach is that the model selection process may result in slightly different terms to a comparable process using multilevel models at every step of selection. However, this is most likely to affect the less influential effects and so, considering resource and time constraints, was found to give a good balance between robustness and feasibility.

borderline significant terms were retested. More details of the model selection process and Stata code used can be found in Appendix 7.

3.6.2 Model results

Table 9 shows the final estimates from the multivariate model of the factors associated with visiting a **library**. The bar chart at the side of the table demonstrates the strength of relationships – red bars for negative effects and green bars for positive effects. The bar sizes are proportional to either the odds ratio (where a variable is binary) or the odds ratio raised to the power of one standard deviation (where a variable is continuous or interval), demonstrating the relative effect of changing a binary variable from 0 to 1, or moving one standard deviation along the scale of a continuous variable

The estimates demonstrated a much more interesting picture than the bivariate analyses, with some very strong effects in evidence. There were strong positive effects on visiting a library for individual cultural influence (whether people feel they have influence over decisions relating to the provision of library services), cycling and being part of a BME group. Moderately strong positive effects were estimated for internet access, education, childhood experience, children living in the household and age. Some quite large negative associations were also present, however. Working full-time and being male were associated with a reduced probability of visiting a library.

Table 9: Multivariate model of the actors associated with visiting a library

Model terms	Odds ratio	Coefficient	Std. Err.	p-value	Less/more likely
Socio-economic characteristics					
Children living in household vs. not	1.3173	0.2756	0.1294	0.033	
Children interaction with living as a couple	1.3176	0.2758	0.1498	0.066	
Living as a couple vs. not	0.9510	-0.0503	0.0780	0.519	
Member of a BME group vs. not	2.5931	0.9529	0.2574	0.000	
Member of a BME group interacted with age	0.9847	-0.0154	0.0060	0.010	
Age	1.0192	0.0190	0.0097	0.050	
Age squared	0.9998	-0.0002	0.0001	0.015	
Sex interaction with age	1.0108	0.0107	0.0030	0.000	
Sex: male vs. female	0.4830	-0.7276	0.1592	0.000	
Income	0.9645	-0.0362	0.0104	0.000	
Religion: non-practicing vs. not religious	0.8984	-0.1071	0.0791	0.176	
Religion: practicing vs. not religious	1.1752	0.1614	0.0893	0.071	
Educated to other higher level vs. degree	0.9818	-0.0184	0.1098	0.867	
Educated to A-level vs. degree	0.8188	-0.1999	0.1230	0.104	
Educated to trade apprentice level vs. degree	0.4222	-0.8623	0.1915	0.000	

Model terms	Odds ratio	Coefficient	Std. Err.	p-value	Less/more likely
Educated to 5+ GCSEs A*-C level vs degree	0.7184	-0.3307	0.1093	0.002	
Educated to <5 GCSEs A*-C level vs degree	0.5306	-0.6338	0.1265	0.000	
Educated other qualifications vs degree	0.4523	-0.7934	0.2225	0.000	
Educated to no qualifications vs. degree	0.4181	-0.8719	0.1047	0.000	
Work status: part time vs. not working	0.8518	-0.1603	0.0943	0.089	
Work status: full time vs. not working	0.4789	-0.7364	0.0839	0.000	
Accessibility of engagement					
Cycles from place to place vs. not	1.8897	0.6364	0.1461	0.000	
Has access to a motor vehicle vs. not	0.8505	-0.1619	0.0849	0.056	
Media access					
Has access to internet vs. not	1.2514	0.2243	0.0787	0.004	
Hours of TV watched per day	0.8887	-0.1180	0.0221	0.000	
Childhood experience					
Encouraged to read as child: a lot vs. not	1.3271	0.2830	0.0828	0.001	
Encouraged to read as child: some vs. not	1.2544	0.2266	0.0967	0.019	
No. of times taken to a library p.a. as child	1.0468	0.0457	0.0056	0.000	
Influence over provision					
Influence over cultural facilities: some vs. none	1.8078	0.5921	0.0934	0.000	
Influence over cultural facilities: lot vs. none	1.5562	0.4422	0.3131	0.158	
Enjoyment of engagement					
% of local population satisfied with libraries	1.0270	0.0266	0.0063	0.000	
constant	-	*-1.6084*	*0.5114*	*0.002*	

number of individual respondents = 10039

number of local authorities = 346

variance of random intercept = 0.0892

standard error of random intercept = 0.0243

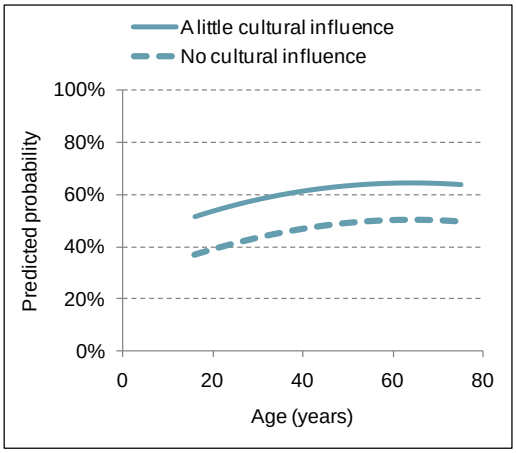
(Note: the bar chart is approximate - the sizes of bars are proportional to the odds ratios raised to the power of one for dummy variables or the odds ratios raised to the power of one standard deviation for continuous variables)

Predictions were created using the model for various combinations of the explanatory variables. *Figure 29a* shows a comparison of the predicted probability of visiting a library by age for individuals who reported cultural influence of “a little” and “none”. The dashed line represents individuals who reported “a little” influence over the provision of library services, and the solid line represents individuals who reported not having any cultural influence.

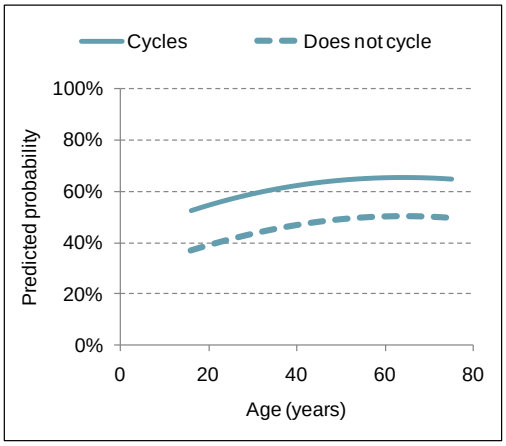
Variations in cultural influence are associated with a difference in the probability of visiting a library between 15% and 20%.

Figure 29: Predicted probability of visiting a library by socio-demographic factors

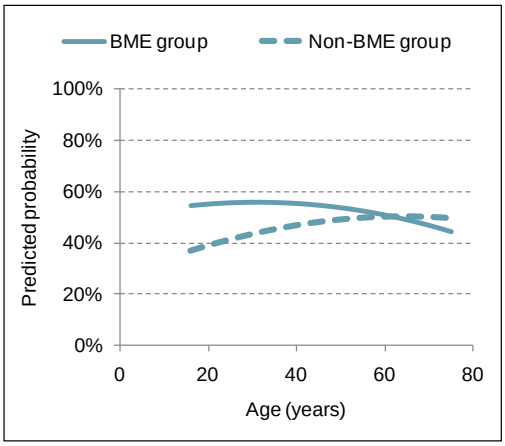
29a. Cultural influence



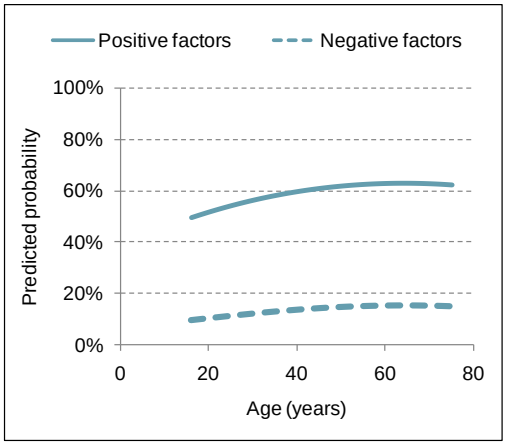
29b. Cycle vs. does not cycle



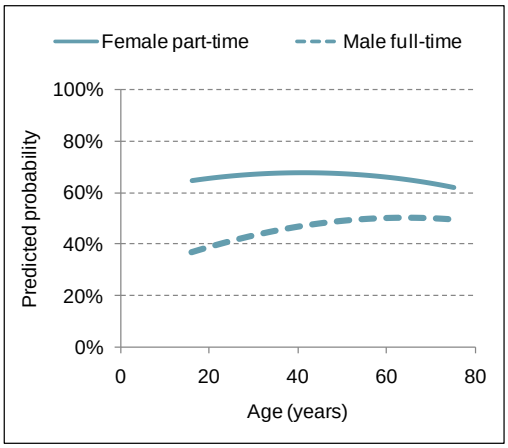
29c. BME groups



29d. Education and childhood



29e. Gender and work patterns



Cycling produced a very similar difference in predicted probability to cultural influence. *Figure 29b* shows the predicted probabilities by age of individuals who cycle with those who

do not. Those who cycle for travel have a 15% and 20% higher probability of visiting a library than those who don't.

Figure 29c shows a comparison of predicted probabilities of visiting a library for BME and non-BME groups. The interaction between BME group and age suggests that the likelihood of those in the BME group visiting a library reduces with age. The opposite trend is observed for non-BME groups.

Moderately strong positive effects were estimated for education, internet access, childhood experience and children living in the household. *Figure 29d* demonstrates the large difference in predicted probability caused by combinations of these factors. The solid line represents a degree-level educated, internet-accessing individual with a very frequent childhood experience of visiting libraries and high childhood parental reading encouragement who also lives in a household with children (a "Positive factors" combination). The dashed line represents an individual with a low level of education (fewer than five A*-C grades at GCSE), no internet access, low frequency childhood experience of visiting libraries and low childhood parental encouragement to read who lives in a household without children (a "Negative factors" combination). Likelihood of visiting a library for the negative combination is approximately a fifth of that for the positive combination across the age range.

Two of the stronger negative effects were related to work status and sex. The predicted probability of visiting a library for individuals who work full-time was lower than for part-timers and the unemployed, and being male was also associated with a reduced probability of library engagement. *Figure 29e* demonstrates these effects combined in a manner maximally detrimental to library engagement. The solid line represents the predicted probability for a female who is a part-time worker, whereas the dashed line represents the predicted probability by age for a male in full-time employment.

3.6.3 Model testing and diagnostics

Software for multilevel modelling is relatively new and is developing as computers become more powerful and are able to cope with the large number of calculations needed to estimate these types of model. As a result, a limited number of fit and diagnostic tests were available for use with weighted multilevel models. In order to test the final library model thoroughly, a single level logistic regression model with robust standard errors (an unweighted version of the model specification used during the selection process, see equation (4) above) was constructed to mimic the multilevel model. Apart from not containing a random intercept parameter, this model contained the same terms as the multilevel library model. In the model, the random intercept term was significant at the 5% level, thus a single-level model does not fit the data as well as the multilevel model. Therefore, if the single-level model was found to fit the data well, using the tests available for single-level unweighted models, it is reasonable to assume that the multilevel model would also fit the data well.

A model specification link test of the single-level logistic regression model did not identify any problems with the model specification (p=0.749) and a Hosmer-Lemeshow test suggested that it was a good fit to the data (50 groups, $Chi^2(48)=42.28$, p=0.7056). A plot of standardised residuals at the local authority level is shown in *Figure 30* and a plot of deviance residuals for the respondent level is shown in *Figure 31*. The plots do not demonstrate any abnormalities that would cause concern. Further information on model testing and diagnostics can be found in Appendix 7.

Figure 30: Local authority level standardised residuals

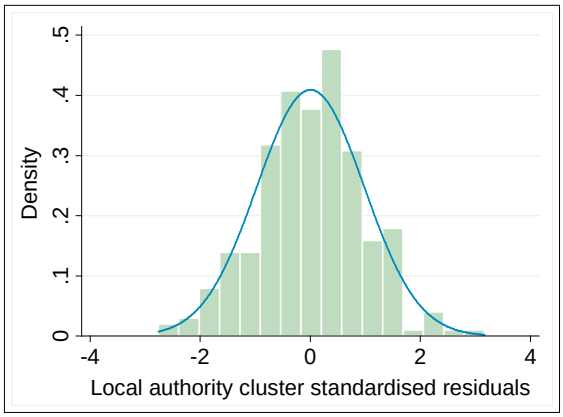
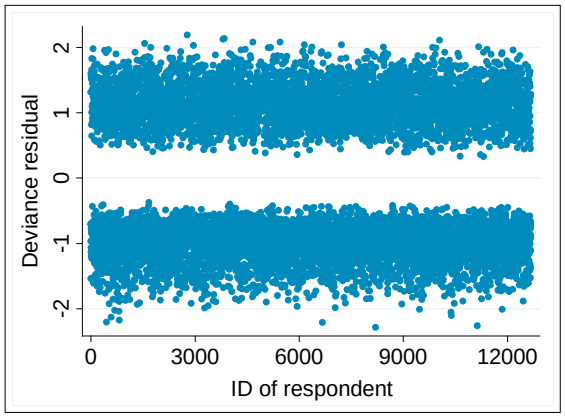


Figure 31: Respondent level deviance residuals



Estimates of explanatory power for logistic regression (known as pseudo-R^2 measures) do not benefit from the straightforward interpretation found in continuous regression models because the response variables can only take the values 0 or 1. As a result, although many different versions are available, statisticians do not view any of these measures as particularly robust and they only serve as weak analogues of the more familiar R^2 in continuous regression. Generally, pseudo-R2 measures return low values of explanatory power of between 10% and 30%. Despite these reservations, two estimates were calculated using the single level robust weighted model to give comparable indications of explanatory power:

- McFadden's R^2 – this resulted in an estimate of a 9.8% reduction in error using the model to predict engagement as compared to the overall mean.
- McKelvey and Zavoina's R^2 – this resulted in an estimate of 16.7% reduction in error using the model to predict engagement as compared to the overall mean.

3.7 Visiting a museum

3.7.1 Model development

An initial analysis was undertaken on the factors associated with visiting a museum, library or archive (MLA). The initial results from this model demonstrated the likely problems arising from modelling a response variable that encompassed engagement with museums, libraries and archives all together. It was therefore decided to separate the component parts of MLA. As the proportion of Taking Part respondents visiting archives was low (approximately 5%), models were run for visiting museums and for visiting libraries.

The museum model was built using the response variable **museum** – whether the respondent had visited a museum in the past 12 months. The model was constructed by populating the conceptual framework (Appendix 1). The variables used to construct this model are described in *Table 10*. More information on the data is available in Appendices 3 and 4.

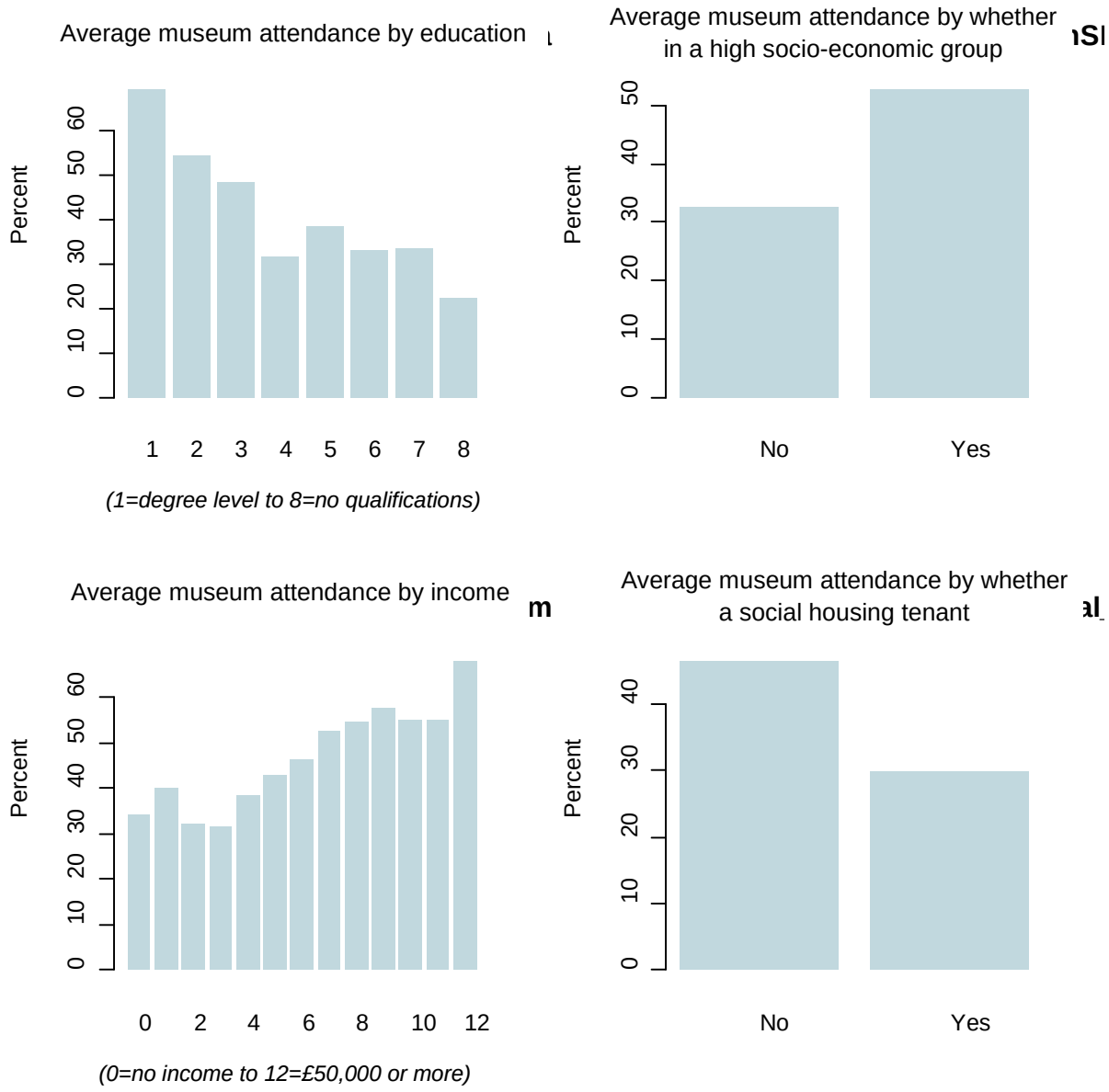
Table 10: Variables used to build museum model

Variable name	Description
access	Accessibility to services index from National Indicator set (percent) – percentage of people who can reach selected core services and facilities within 15 minutes by public transport and/or walking (National Indicator 175)
age	Age in years
BMEgroup	Member of ethnic minority group (1=yes, 0=no)
child_heritage_visit	Average times per year taken to heritage sites when growing up
child_museum_visit	Average times per year taken to museums or art galleries when growing up
children	Children living in household (1=yes, 0=no)
community_cohesion	% of people who feel that they belong to their neighbourhood
coupled	Living as part of a couple (1=yes, 0=no)
cultural_influence	Has influence over local cultural facilities (categorical from 0=no to 2=a lot)
cycles	Cycles to get from place to place (1=yes, 0=no)
education	Highest educational qualification held (categorical from 1=degree to 8=no qualifications)
highSES	Member of high socio-economic group NS-SEC 1-4 (1=yes, 0=no)
history_tv	Watches historical television programmes (1=yes, 0=no)
income	Highest income in household (interval from 0=£0 to 12=£50,000 or more)
internet	Has access to internet (1=yes, 0=no)
limiting_illness	Has illness or disability which limits activities (1=yes, 0=no)
local_museum_satisfaction	% of people who are very or fairly satisfied with museums/galleries
local_museums	Population weighted count of museums in area – categorised by SIC(92) classification code 9252 (Museum Activities And Preservation Of Historical Sites And Buildings)
motor_vehicle	Has access to a motor vehicle (1=yes, 0=no)
museum	Visited museum in last 12 months (1=yes, 0=no)
museum_internet	Visited museum website in last 12 months (1=yes, 0=no)
newspaper	Reads daily newspaper at least three times per week (1=yes, 0=no)
radio	Radio available in household (1=yes, 0=no)
religious	Religiosity (0=not religious, 1=non-practising, 2=practising)
science_tv	Watches science television programmes (1=yes, 0=no)
sex	Sex of respondent (1=male, 0=female)
social_housing	Is a social housing tenant (1=yes, 0=no)
tv_hours	Hours of television watched per day on average
work_status	Work status (0=not working, 1=part-time, 2=full-time)

Bivariate analyses of the explanatory variables with **museum** identified many significant associations (see Appendix 5). Some of the strongest associations involved education,

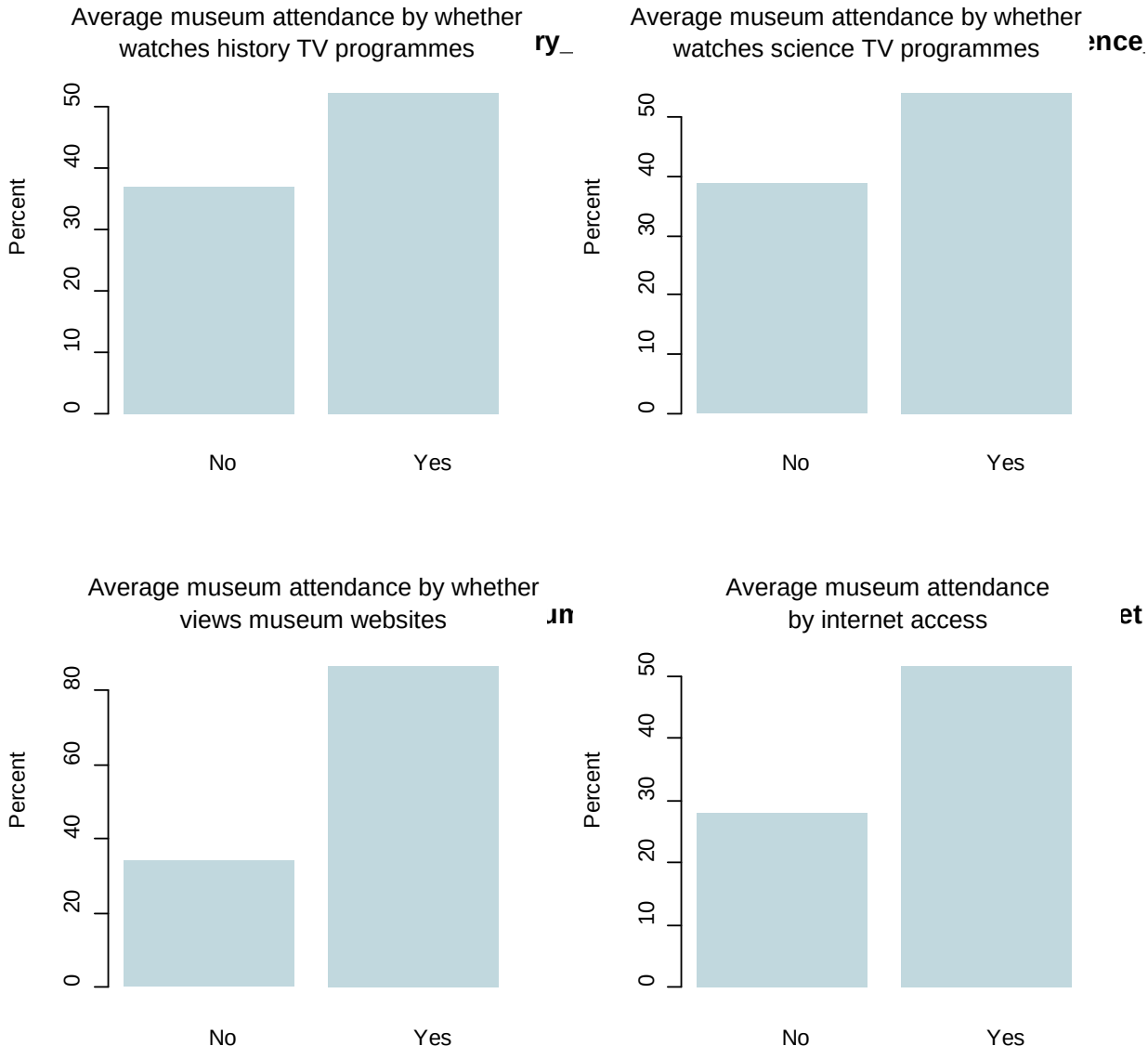
socio-economic, media and childhood experience factors. *Figure 32* overleaf shows plots of the likelihood of visiting a museum by education, socio-economic status, income and whether the individual lives in social housing. The likelihood of visiting a museum is lower for social housing tenants. Increasing levels of education (category 1 indicates degree-level educated), increasing income and high socio-economic status are all related to an increased likelihood of visiting a museum.

Figure 32: Likelihood of visiting a museum by education, socio-economic status, income and social housing tenure



The likelihood of visiting a museum was positive associated with certain types of media consumption. *Figure 33* demonstrates that watching history and science-related TV programmes and having access to the internet and visiting museum websites were associated with a higher probability of visiting a museum.

Figure 33: LIkelihod of visiting a museum by television and internet media consumption



The likelihood of visiting a museum was, however, negatively associated with hours of TV viewing. *Figure 34* shows the relationship between the likelihood of visiting a museum and hours of TV watched per night.

Figure 34: Average museum engagement by average hours of television watched per night

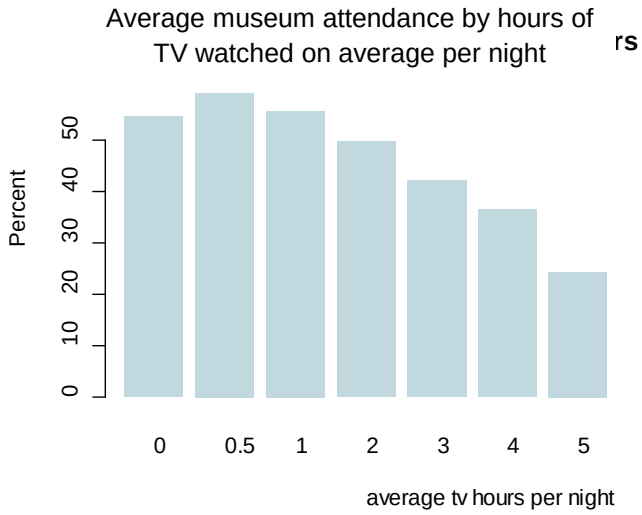
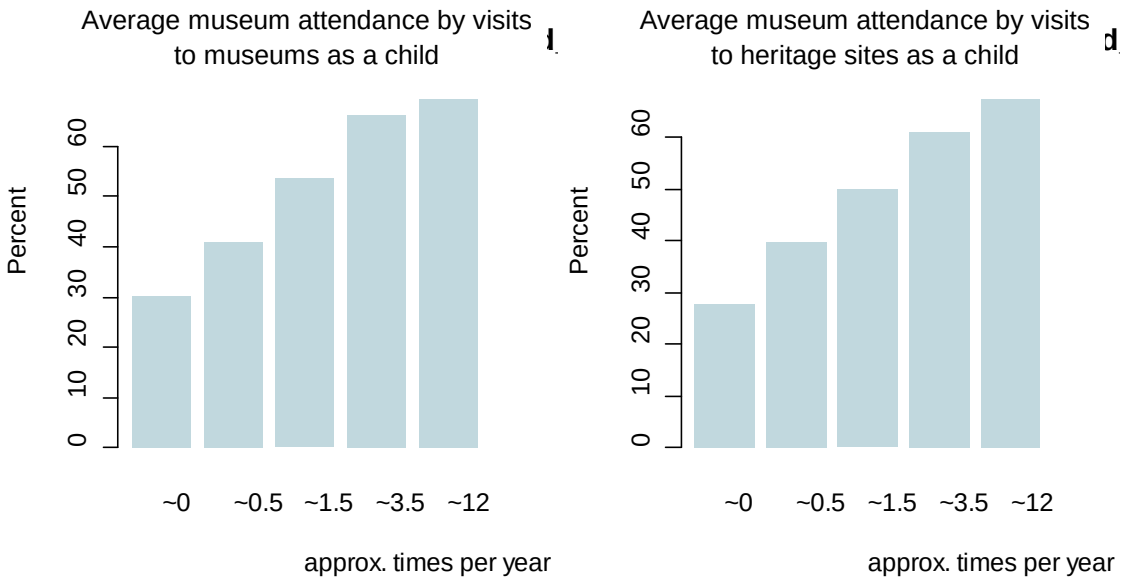


Figure 35 demonstrates that childhood experience of visiting museums was positively associated with the likelihood of visiting a museum as an adult.

Figure 35: Likelihood of visiting a museum by childhood experience of visiting museums and heritage sites



Interval and ratio variables were plotted to determine whether any exhibited a high degree of skew, which could lead to instability in model estimation (see Appendix 4). The original variable **local_museums** was intended as a measure of supply and was simply a count of museums by local authority – it did not take account of the population in those authorities. In order to take account of local authority population size, producing a measure of supply density, the data were weighted by 2007 mid-year ONS population estimates (in thousands). The variable was highly skewed and so it was log transformed ($\log_e(X+0.01)$) in order to reduce skew.

Previous research on the drivers of engagement has suggested that multicolinearity among explanatory variables (e.g. **income** and **education**) leads to difficulties in isolating their effects on engagement (see section 3). In order to test whether multicolinearity was likely to be a problem in our data, variance inflation factors (VIF) were calculated (see Appendix 6). None of the VIFs indicated that problematic multicolinearity was present in the explanatory variables (based on a rule of thumb that VIFs should not exceed 10).

Multivariate model selection was undertaken using robust single level logistic regression of the form[9]:

$$\begin{aligned} L_{ij} &= Logit\left\{\Pr(y_{ij}=1 \middle| x_{1j}, x_{2j}, x_{3ij}, x_{4ij}, ..., x_{nij})\right\} \\ &= \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + ... + \beta_n x_{nij} \end{aligned} \quad (5)$$

Where:

$i =$	Identifier for individual
$j =$	Identifier for area
$L_{ij} =$	Natural logarithm of odds of 'yes' response to a yes/no question
$y_{ij} =$	Response to yes/no question as a Boolean value
$\beta_0, \beta_1, \beta_2 ... =$	Coefficients of regression model
$x_{1j}, x_{2j} ... =$	Explanatory variables at area level
$x_{3ij}, x_{4ij}, ..., x_{nij} =$	Explanatory variables at individual level

The data were weighted using the variable **pweight1** (individual level weight, see Appendix 7 for more information on data). Model selection proceeded by removing terms in order of decreasing significance. A nominal significance level of 5% was used to determine whether terms were retained in the model. Once the terms remaining were all significant at the 5% level, interactions to identify additional variations in effect due to demographics, family structure, socio-economic group and free time were added. Finally, the model was re-

[9] Model selection was undertaken using robust single level logistic regression rather than the multilevel specification shown in equations (1) & (3). The risk to this two stage approach is that the model selection process may result in slightly different terms to a comparable process using multilevel models at every step of selection. However, this is most likely to affect the less influential effects and so, considering resource and time constraints, was found to give a good balance between robustness and feasibility.

estimated with the final set of terms using the specification in equations (1) & (3) and weighted with **pweight1** and **pweight2** (area level weight).

During the model selection process, the effect of viewing museum websites (**museum_internet**) was identified as problematic. The effect size was much greater than for any other covariate. The likely reason for this very large effect was ambiguity in the causal direction, as visiting museum websites might result from visiting a museum, rather than be a precursor to museum visiting. Therefore, **museum_internet** was removed from the model, and some previously borderline significant terms were re-tested. More details of the model selection process and Stata code used can be found in Appendix 7.

3.7.2 Model results

Table 11 shows the final estimates from the multivariate modelling of whether people visit a **museum**. The bar chart at the side of the table demonstrates the strength of relationships – red bars for negative effects and green bars for positive effects. The bar sizes are proportional to either the odds ratio (where a variable is binary) or the odds ratio raised to the power of one standard deviation (where a variable is continuous or interval), demonstrating the relative effect of changing a binary variable from 0 to 1, or moving one standard deviation along the scale of a continuous variable

The estimates demonstrated strong positive effects on the likelihood of visiting a museum of media consumption, childhood experience, socio-economic factors and cultural influence. The likelihood of visiting a museum was negatively associated with being part of a BME group, being part of a couple and being male. The probability of visiting a museum tended to rise with increasing level of education.

Table 11: Multivariate model of the factors associated with visiting a museum

Model terms	Odds ratio	Coefficient	Std. Err.	p-value	Less/more likely
Socio-economic characteristics					
Age	1.0414	0.0405	0.0104	0.000	
Age squared	0.9995	-0.0005	0.0001	0.000	
Member of a BME group vs. not	1.1444	0.1349	0.4018	0.737	
Member of a BME group interacted with age	0.9798	-0.0204	0.0098	0.036	
Living as a couple vs. not	0.5109	-0.6715	0.1958	0.001	
Living as a couple interaction with age	1.0101	0.0100	0.0033	0.002	
Sex interaction with living as a couple	1.5812	0.4582	0.1370	0.001	
Sex: male vs. female	0.6648	-0.4082	0.0996	0.000	
High social-economic status vs. not	1.1807	0.1661	0.0665	0.013	
Income	1.0326	0.0321	0.0117	0.006	
Religion: non-practicing vs. not religious	0.8337	-0.1819	0.0879	0.039	
Religion: practicing vs. not religious	1.0792	0.0762	0.0990	0.442	

Model terms	Odds ratio	Coefficient	Std. Err.	p-value	Less/more likely
Work status: part time vs. not working	0.8774	-0.1308	0.1014	0.197	
Work status: full time vs. not working	0.6658	-0.4068	0.0774	0.000	
Illness or disability vs. not	0.7372	-0.3048	0.0794	0.000	
Educated to other higher level vs. degree	0.6317	-0.4594	0.1354	0.001	
Educated to A-level vs. degree	0.6255	-0.4692	0.1106	0.000	
Educated to 5+ GCSEs A*-C level vs degree	0.4565	-0.7842	0.1129	0.000	
Educated to <5 GCSEs A*-C level vs degree	0.3807	-0.9656	0.1350	0.000	
Educated other qualifications vs degree	0.3738	-0.9840	0.2084	0.000	
Educated to no qualifications vs. degree	0.3440	-1.0671	0.1181	0.000	
Educated to trade apprentice level vs. degree	0.3067	-1.1818	0.2020	0.000	
Accessibility of engagement					
Index of service accessibility	1.0160	0.0159	0.0049	0.001	
Has access to a motor vehicle vs. not	0.8322	-0.1836	0.0924	0.047	
Media access					
Watches history on TV vs. not	1.7235	0.5443	0.0684	0.000	
Radio available in house vs. not	1.5095	0.4118	0.1605	0.010	
Has access to internet vs. not	1.4602	0.3786	0.0805	0.000	
Watches science on TV vs. not	1.2099	0.1905	0.0775	0.014	
Hours of TV watched per day	0.8402	-0.1741	0.0269	0.000	
Childhood experience					
No. of times taken to museums p.a. as child	1.1177	0.1112	0.0160	0.000	
No. of times taken to heritage sites p.a. as child	1.0628	0.0609	0.0132	0.000	
Influence over provision					
Influence over cultural facilities: some vs. none	1.7291	0.5476	0.0997	0.000	
Influence over cultural facilities: lot vs. none	1.2522	0.2249	0.3018	0.456	
constant	-	*-2.2219*	*0.5506*	*0.000*	

number of individual respondents = 9903

number of local authorities = 346

variance of random intercept = 0.1424

standard error of random intercept = 0.0304

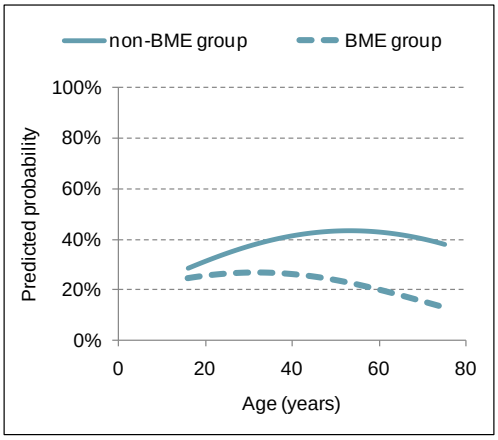
(Note: the bar chart is approximate - the sizes of bars are proportional to the odds ratios raised to the power of one for dummy variables or the odds ratios raised to the power of one standard deviation for continuous variables)

Predictions were created using the model for various combinations of the explanatory variables. *Figure 36a* shows a comparison of the predicted probability of visiting a museum by age for individuals who are or are not part of a BME group. The dashed line represents individuals in a BME group, and the solid line represents those not in a BME group. It demonstrates that the difference in likelihood of visiting a museum between BME and non-BME groups increases with age. That is, while the two groups have a similar probability of visiting a museum at 20 years old, by the time they are 60 years old, the probability of visiting a museum is 20% higher for non-BME groups.

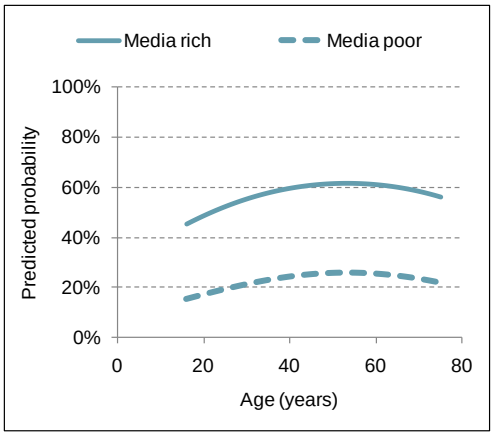
Combined differences in media consumption resulted in large changes in predicted probability. *Figure 36b* shows a comparison of individuals with internet access, who watch history and science TV programmes and have a radio in the household (the "media-rich" – solid line) compared with those who do not have internet access, do not watch history and science TV programmes and do not have a radio (the "media-poor" – dashed line). It demonstrates that the probability of visiting a museum is about 30% higher for the media-rich.

Figure 36: Predicted probability of visiting a museum by socio-demographic factors

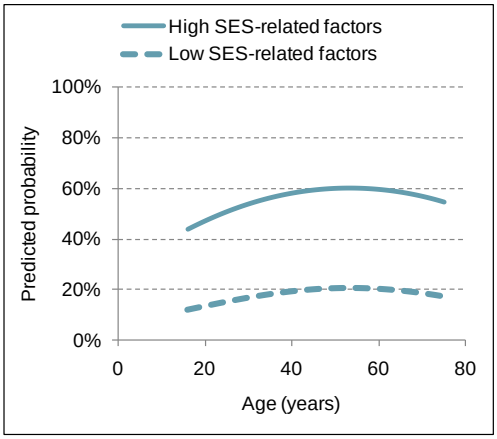
36a. BME groups



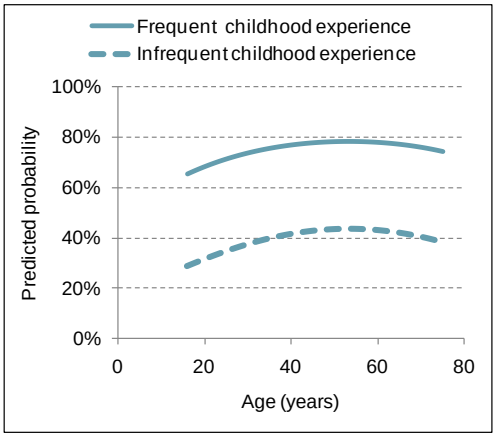
36b. Media consumption



36c. High vs. low SES



36d. Childhood experience



Combined differences in socio-economically related factors resulted in large changes in predicted probability of visiting a museum. *Figure 36c* shows a comparison of the predicted probability of visiting a museum by age for individuals who are high on all factors related to socio-economic advantage with those who are low on the same factors. The dashed line represents individuals who are of low socio-economic status, have an annual income of £10,000 to £14,999, who watch four hours of television per night and have an education level of fewer than five A*-C grades at GCSE. The solid line represents individuals who are of high socio-economic status, have an annual income of £50,000 or more, who watch two hours of television per night and have an education level of degree or above. Combined together, these factors have a very large impact on the predicted probability of visiting a museum, with 50% to 60% of high socio-economic individuals predicted to visit compared with less than 20% of low socio-economic individuals.

There were also strong positive effects related to childhood experience of museum and heritage engagement. *Figure 36d* shows the predicted probability of visiting a museum for individuals who visited museums and heritage sites 10 times a year on average when they were children (solid line) compared with those who visited only once a year on average (dashed line). The predicted probability of visiting a museum as an adult for the individuals with infrequent childhood visits is approximately half that for those with frequent childhood visits.

3.7.3 Model testing and diagnostics

Software for multilevel modelling is relatively new and is developing as computers become more powerful and are able to cope with the large number of calculations needed to estimate these types of model. As a result, a limited number of fit and diagnostic tests were available for use with weighted multilevel models. In order to test the final museum model thoroughly, a single level logistic regression model with robust standard errors (an unweighted version of the model specification used during the selection process, see equation (5)) was constructed to mimic the multilevel model. Apart from not containing a random intercept parameter, this model contained the same terms as the multilevel museum model. In the model, the random intercept term was significant at the 5% level, thus a single-level model does not fit the data as well as the multilevel model. Therefore, if the single-level model was found to fit the data well, using the tests available for single-level unweighted models, it is reasonable to assume that the multilevel model would also fit the data well.

A model specification link test of the single-level logistic regression model did not identify any problems with the model specification (p=0.247) and a Hosmer-Lemeshow test suggested that it was a good fit to the data (50 groups, $Chi^2(48)=42.28$, p=0.7056). A plot of standardised residuals at the local authority level is shown in *Figure 37* and a plot of deviance residuals for the respondent level is shown in *Figure 38*. The plots do not demonstrate any abnormalities that would cause concern. Further information on model testing and diagnostics can be found in Appendix 7.

Figure 37: Local authority level standardised residuals

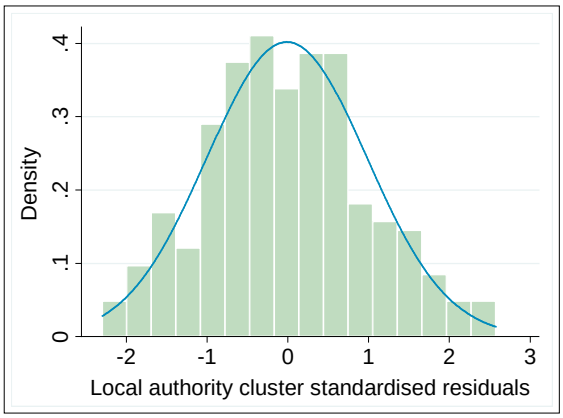
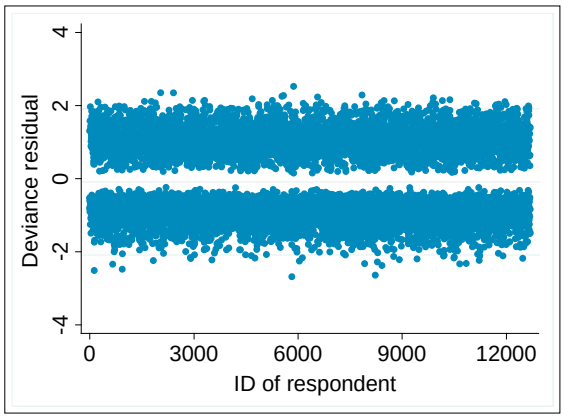


Figure 38: Respondent level deviance residuals



Estimates of explanatory power for logistic regression (known as pseudo-R^2 measures) do not benefit from the straightforward interpretation found in continuous regression models because the response variables can only take the values 0 or 1. As a result, although many different versions are available, statisticians do not view any of these measures as particularly robust and they only serve as weak analogues of the more familiar R^2 in continuous regression. Generally, pseudo-R2 measures return low values of explanatory power of between 10% and 30%. Despite these reservations, two estimates were calculated using the single level robust weighted model to give comparable indications of explanatory power:

- McFadden's R^2 – this resulted in an estimate of a 15.2% reduction in error using the model to predict engagement as compared to the overall mean.
- McKelvey and Zavoina's R^2 – this resulted in an estimate of 25.9% reduction in error using the model to predict engagement as compared to the overall mean.

3.8 Discussion

A number of trends emerge across the models of different engagement types, including:

- Older people are more likely to engage in culture, but less likely to engage in sport.
- Members of BME groups are more likely to engage in culture, but less likely to engage in sport. This effect is particularly strong for libraries, with members of BME groups being much more likely to visit a library than non-BME groups.
- Childhood experience of engaging in all types of culture is positively associated with engaging in culture as an adult.
- Those with higher levels of education are more likely to engage in culture.
- Those of higher social economic status are more likely to attend arts events, visit a heritage site, and visit a museum.
- Media consumption is positively associated with engagement in culture and sport. For instance, having access to the internet is positive associated with engaging in culture.

- Men are much more likely to participate in sport, but less likely to attend arts events, visit a museum, or visit a library.

The effect of BME group status on the probability of engagement in culture varies with age. That is, in the cases of visiting a heritage site, attending an arts event, and visiting a museum, young people from BME and non-BME groups have a similar probability of engaging in culture, while among older people those from a BME group are less likely to engage in culture. The reverse of this trend is observed for visits to libraries. The greater likelihood of younger generations of BME groups engaging in culture may have important implications for ensuring social cohesion.

There is a positive association between whether people watch culture and sport related TV programmes and whether they engage in culture and sport. It is, however, likely that this association is explained by an underlying interest in culture and sport, rather than TV watching having a causal effect on actual attendance of cultural events / site or participation in sport.

A number of interesting engagement type-specific trends also emerge from the analysis, including:

- Those people with access to a motor vehicle are more likely to visit a heritage site.
- Those people with a limiting illness are less likely to participate in sport.
- Those people who cycle are more likely to visit a library.
- Those people who drink alcohol are more likely to attend art events.

The first two of these observations make intuitive sense and serve to validate the results of the analysis. The observation that people who cycle are more likely to visit a library is probably explained by the fact that both factors are associated with a certain type of lifetime, rather than having any particular policy implication. The observation that people who drink alcohol are more likely to attend arts events is probably explained by arts events being social occasions.

Section 3 identified two traditions in the extant literature on the drivers of engagement in culture and sport. First, the econometric approach informed by the general theory of consumer choice emphasises the importance of price and income in driving engagement. Second, the broader social science literature emphasises a broader set of socio-demographic factors, allowing analysis of, for instance, the impact of available time and the factors that influence preference development, such as education and previous experience. The analysis reported in this section confirms the importance of the socio-demographic factors emphasised by the broader social science literature.

Income had an effect in each of the models, but this effect was small and the direction of the effect varied between engagement types. A similarly small effect of income on demand is identified in the econometric literature on the demand for attendance at performing arts events (see section 3). One explanation forwarded for this observation was the failure of extant studies to adequately account for the substitution effect of increased income – the possibility that people participate in fewer leisure activities as the opportunity cost of these activities (the value of work) increases. This hypothesis is supported by the analysis of

Taking Part reporting in section three demonstrating a positive association between income and the proportion of respondents who report limited time as a barrier to engagement. The analysis reported above controlled for working status, but this is a blunt measure of the amount of the distribution of time between work and leisure.

Section 3 discussed the role of price in influencing demand for engagement. The datasets available did not provide an estimate of the price of engagement, so it is not possible to assess the relationship between price and engagement. However, there are a number of reasons why the analysis would not have identified a relationship between such a variable and engagement should it have been included (see section 3 for further detail). First, it is possible that the price of engagement makes up only a small proportion of the cost of engagement. Other costs, in particular the opportunity cost of the time required to engage in culture and sport, perhaps make up a greater proportion of the cost of engagement.

Second, given the aggregate nature of the analysis, including price in the analysis would not have been meaningful. One of the explanations posited for the inconclusive nature of econometric attempts to measure the price elasticity of demand for attending a performing arts event was that this analysis was undertaken at too aggregate a level. That is, the analysis was insufficiently focused on a particular type of performing art. Thus, even if it was possible to construct a general price variable, the heterogeneity in the activities included in the engagement variables, and thus the variation in the relationship between price and engagement, mean that it would have been unlikely that an effect of price would have been observed.

Third, very few people report affordability as a barrier to engagement. Table 12 shows an analysis of the Taking Part survey that demonstrates only a small proportion of non-engagers report affordability as the reason they don't engage.

Table 12: Proportion of non-engagers who mention limited affordability as a barrier to engaging in culture and sport (Taking Part, 2006/7)

	Income level		
	Low	Average	High
Attending arts events	7%	5%	5%
Attending a heritage site	5%	2%	4%
Attending a library	4%	2%	1%
Attending a museum	3%	2%	1%
Participating in sport	2%	3%	3%

Implications for policy making

One of the criticisms of the existing literature is its inability to provide policy relevant insights into the drivers of engagement in culture and sport (see section 3). Which of the policy levers at the disposal of DCMS and its related non-departmental public bodies are most effective at improving engagement? An important policy lever is the price of entry/attendance. However, as noted above, it was not possible, nor would it have been meaningful within the analytical

approach adopted, to include a variable for price. What other policy levers does the analysis talk to?

An important lever for policy makers is the provision of quality assets/sites for engagement. For instance, what effect does the restoration of heritage sites have on the probability that people visit heritage sites? With the exception of heritage assets, the analysis found no association between the supply of sites/ assets and the probability that people engage in culture and sport. The importance of the local supply of heritage assets in determining the probability of visiting a heritage site is supported by the observation that access to a motor vehicle also influences the probability of visiting a heritage site. This observation is not surprising given the geographic distribution of heritage sites. A negative association was observed, however, between access to a motor vehicle and the probability of visiting a library or museum. Rather than suggesting that access and/or supply is negatively associated with visiting museums and/or libraries, this observation might be picking up the influence of whether people live in urban or rural locations.

Similarly, with the exception of libraries, the analysis found no association between the quality of sites/assets and the probability that people engage in culture and sport. The latter observation may suggest that, of the engagement types considered, quality of facilities is only important in determining engagement in libraries, or that there is greater variation in the quality of libraries. The latter possibility is not supported by the data, as satisfaction with libraries has a relatively low level of variation when compared with other engagement types (see Appendix 4).

It is important to note however, that these observations that there is a lack of relationship between the supply and quality of assets and the probability of engaging in culture and sport could be explained by the weaknesses of the area-level variables used to measure supply and quality. Further discussion of these variables is available later in this section.

Related to the issue of supply is that of accessibility. The analysis found a positive association between the index of service accessibility and the probability of visiting a museum, but a negative association with the probability of visiting a heritage site. It is possible that this observation can be explained by the accessibility index acting as proxy for urban-rural distinctions. That is, to the extent that service accessibility is higher in urban areas, it might be picking up the greater the proximity of museums.

An issue related to the supply and quality of assets/sites is that of disabled access. The analysis identified a negative association between having a limiting illness and the probability of doing sport, visiting a museum and attending an art event. This might imply that improvements in disabled access and outreach activities for museums and art events could increase engagement. It is important to note, however, that the measure of limiting illness used in the analysis includes conditions whose impact on the possibility of engagement is unlikely to be mitigated by policies on disabled access. That is, it is not possible to alleviate the effect of all illness that limit peoples' ability to engage in culture and sport through public policy.

Another policy implication of the research is the idea that allowing people greater influence over decision-making might improve engagement. A positive association was identified

between the extent of influence and the probability of visiting a library or a museum, attending an art event, and doing sport. It is possible, however, that the line of causation runs from engagement to influence. That is, those people who engage more are perhaps also more likely to get involved in the running of their local club or cultural asset. Such involvement may thus be an indication of engagement, rather than a cause of such engagement.

The positive association between media access and the probability of engagement might have implications for policies to increase awareness of engagement possibilities. There was a positive association between whether people had access to the internet or radio and the probability that they would engage in culture and sport. There are a number of possible interpretations of this observation. First, access to the media could be a proxy for information on engagement opportunities. Second access to the media could be a proxy of socio-economic factors. The fact that a range of socio-economic factors are controlled for in the models suggests that the former interpretation might be more appropriate. This in turn would suggest that policies to improve awareness of cultural and sporting opportunities could increase engagement.

Implications for research

The modelling approach adopted was designed to provide a number of advantages over the analysis reported in the existing literature (see section 3 for a discussion of the existing literature). First, consistency across engagement types – the models used common data sources and specifications. Therefore, the results are more comparable across the five engagement types modelled, and discussion of the relative effect of explanatory variables on different engagement types is thus more valid. Previous research has tended to focus on a single engagement type, resulting in considerable variation in data sources, methods, rigour and results.

Second, the employment of multiple datasets and sophisticated model specification – the matching of different datasets and the use of multilevel models allow inclusion of area-level variables (such as number of sporting facilities in an area) and a degree of area-dependent heterogeneity. Studies in the existing literature have tended to analyse the data available in single datasets. Furthermore, to the best of our knowledge, this is the first time multilevel regression models have been used to analyse the Taking Part survey data.

Despite these advantages, inevitably given the complexity of subject matter, there are a number of limitations with the analysis reported in this section, and a corresponding requirement for more research. A number of these limitations result from the inevitable need to focus the analysis, including:

1. Models for other engagement types – the research focused on modelling aggregated measures of engagement. For instance, doing any sport included in the Sport England “1 million” indicator. Further research could model engagement in different types of sport or culture separately.
2. Analysis of changes over time – the analysis used the latest tranche of the Taking Part survey data available to the team (2007/8 survey data). The Taking Part survey has been running since 2005/06 and is ongoing. There have been changes to

questions and sampling strategy during this period, but it might be possible to run similar models for each of these time points to see what changes occur in the model estimates, therefore testing the stability of the relationships over time.

3. Models of engagement frequency – all of the models presented in this paper are logistic regression models of binary response variables (e.g. visited heritage site in past 12 months: yes/no), which can be used to predict the probability of a person engaging in culture or sport. Alternative specifications (such as Poisson models) or modelling approaches (such as event history analysis) are available that enable the frequency of engagement to be modelled.

A number of limitations arise from theoretical questions about the factors that influence engagement in culture and sport. For instance, education has featured as an important explanatory variable in each of the models built. Despite this, research on engagement does not offer an explanation for this relationship. What is it about education that encourages engagement? For instance, if education has a direct causal effect on the probability of engagement, this extra value associated with education would have important policy implications. Such a finding might have a theoretical basis in the idea that education increases exposure to certain types of people, changing norms and preferences (Harris, 1999; and Christakis and Fowler, 2009). If, however, both higher levels of education and higher engagement in culture both reflected an underlying attitude to education, simply ensuring higher levels of education may not have the anticipated effect on engagement levels unless they are accompanied by changes in attitude to learning.

A third group of limitations relates to the nature and quality of the existing data. Despite reviewing and drawing on multiple datasets to construct the analysis, there were still limitations in the available data. This is particularly the case from the perspective of econometric approaches to analysing the drivers of demand for culture and sport. Measures of two key variables required for an econometric analysis – income and price – are limited. As already noted, there is no price data available at the level the analysis was undertaken. Furthermore, the measures of income available in Taking Part are quite blunt, being reported in bands.

The measures of childhood encouragement and experience in the Taking Part survey were found to be strongly associated with engagement in our models. There is some debate as to what the questions really measure, however (Oskala et al. 2009). In order to clarify whether childhood experience is as important as it seems, improvements could be made in methods of measurement, and childhood participation levels could be included in a longitudinal study design. For instance, Oskala et al. (2009) argued that the Taking Part adult survey only collects broad indicators of childhood arts experiences within the family, and it cannot answer all policy questions. They point to the relative impact of family, school, and other agencies at various stages in people's lives as a key area for further research. In particular, as most policies aimed at children operate through the state school system, a better understanding of the role that educational institutions can play in influencing engagement would be of particular policy importance,

Another issue not addressed by the analysis is cross-sector influences. That is, how engagement in one sector influences engagement in another. As discussed in section 3, this is a question inadequately addressed in the existing literature. To what extent can

engagement in different types of culture and sport be considered complements of substitutes for one another? Further research is required to explore this question.

Currently, the models do not take sufficient account of the environment in which people live and work. A significant improvement to the models would be to use geographic information system (GIS) software to incorporate more accurate measures of asset proximity and accessibility. Brook (2005) has built models for art engagement using GIS data that include measures of geographic accessibility both in terms of a person's home location and his or her work location. It should be possible to use similar methods to create a measure that indicates the number and ease of access to assets for each individual in the Taking Part data.

Generally, the Taking Part data is very robust and can be relied on to provide accurate information on the intended factors. Limitations in other data available to the analysis, however, meant that hypotheses could not always be tested. For instance, our model specifications also included local authority-level estimates from other data sources for factors such as asset availability, asset quality, transport infrastructure, and community cohesion. These estimates were at the local authority level because data at a lower level were not available. Also, the sample sizes available from the Taking Part survey at levels lower than local authority were insufficient for robust estimation of area-level effects. The disadvantage of this approach is that lower-level variation is effectively ignored. Another important limitation was the lack of data on the cost of accessing assets.

Finally, as discussed in section 3, the regression analysis framework is subject to a number of limitations. In particular, as illustrated by the previous paragraphs, it is limited by the available data. Furthermore, it is difficult within the regression analysis framework to accurately capture the structure of the decision to engage in culture and sport. Other modelling techniques, such as system dynamics, provide alternative ways to model this decision while drawing on a wider range of data. These techniques provide the possibility of more accurately operationalising a theory of engagement, enabling the formulation and testing of more refined hypotheses than can be achieved with regression analysis. The next section presents the results of a system dynamics model of the engagement in culture and sport.

4. The factors that impact engagement: A system dynamics model

4.1 Introduction

What drives engagement in culture and sport? The previous section summarised the results of a regression analysis of existing survey data designed to answer this question. It also noted a number of the limitations of using this analytical approach to answer this question. First, it requires large survey and/or administrative datasets. Any driver of engagement not included in these datasets is, thus, excluded from the analysis. Second, the structure of the analysis fails to reflect the structure of the problem. While econometric analysis is explicitly built on economic theory, the structure of real world decision making is difficult to capture within the analytical structure and the data requirements of these techniques. This section presents a system dynamics modelling approach designed to overcome these challenges

Rather than there being one decision, engagement in culture and sport requires a number of potentially sequential decisions, each one influenced by a different set of factors. For instance, the price of a ticket to attend an art event is only relevant once a person has decided to attend (McCarthy et al., 2001). The system-dynamic model summarised in this section is designed to capture the sequential nature of different decisions before someone engages in culture or sport.

The previous section outlined some of the factors thought to impact on engagement in culture and sport that are not measured in existing survey datasets. System dynamics models are able to draw on a range of data other than that available in large administrative or survey datasets. Thus, the system dynamics model summarised in this section is also designed to overcome some of the data limitations of the more conventional regression analysis approaches to the problem.

The next section summarises the system dynamics model and the methods employed to design, construct, populate, and test it. The following section presents the results of the application of the model to answer two questions:

1. What is the likely effect of different policy levers on engagement in culture and sport?
2. What is the likely effect of future socio-economic trends on engagement in culture and sport?

The final section summarises the findings of the analysis, considers its policy implications, and reflects on the role of system dynamics in the analysis of complex policy problems such as modelling the effect of policy on engagement in culture and sport.

The system dynamics model reported in this section will be made available in a form that can be used by stakeholders to assess the likely effects of policies and socio-economic trends on engagement in culture and sport. A users' manual will be developed to accompany the model. This report is, thus, not intended to address the use of the model. However, a number of case studies are included to illustrate how the model will be used.

4.2 Method

4.2.1 Overview of method

Figure 39 shows the steps involved in the construction of the system dynamics model of engagement in culture and sport.

Figure 39: Steps in the constuction of the system dynamics model

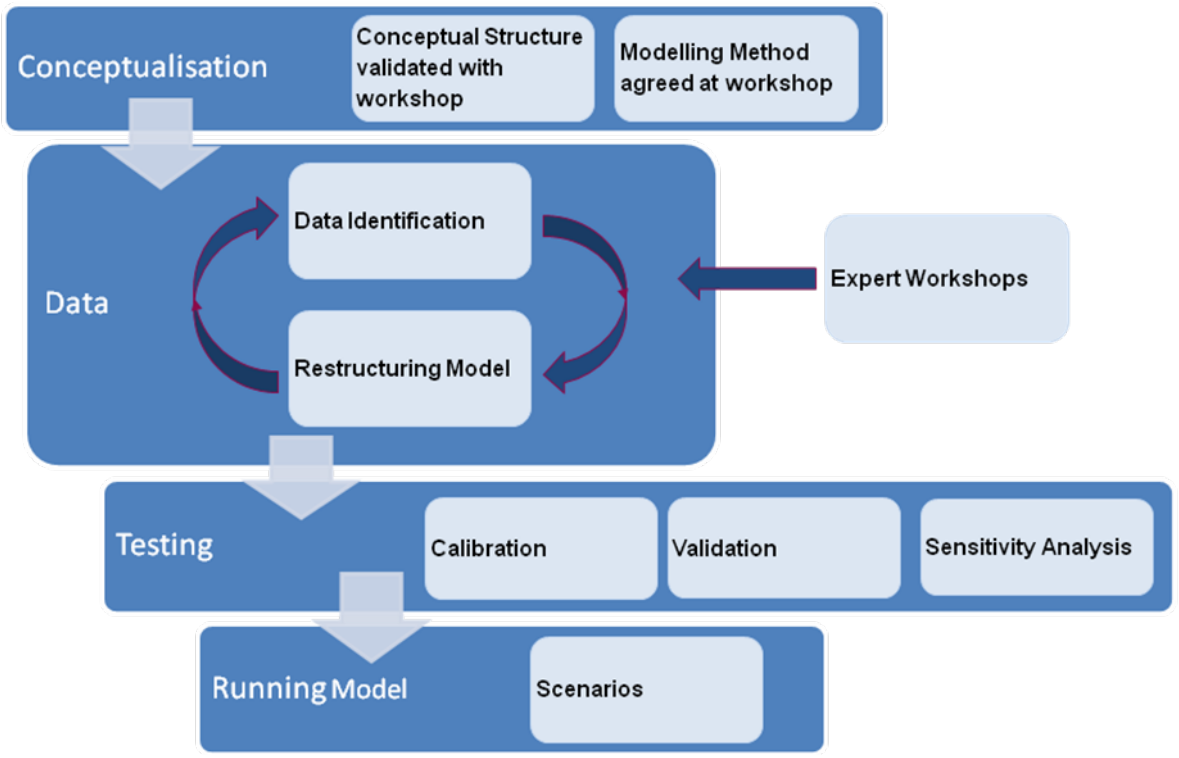


Figure 39 illustrates iterations between model specification and data collection necessary to arrive at the final model. This method section and the following results section report the final model.

4.2.2 Model structure

System dynamics models were developed for the following engagement types:

- Heritage – visiting a heritage site in the past 12 months[10]
- Art – attending an arts event in the past 12 months.
- Sport – whether a person has done three episodes of at least 30 minutes of moderate-intensity sporting activity in the past four weeks (as defined in the Sport England “1 million” indicator).

[10] Based on the Public Service Agreement target for heritage visits by priority groups set up under the Labour administration.

- Museum, library, and archives – whether a person has visited a museum, a library or an archive in the past 12 months.

Within each of these engagement types, separate models were specified for a range of activity types. *Table 13* summarises the activity types for which models were specified. These were chosen as they were the highest-volume activities based on the 2007/2008 Taking Part survey.

A conceptual model of the factors that drive engagement in culture and sport, as well as how these factors interact to generate engagement, was developed by drawing on the extant literature (see section 3). The conceptual model was validated and refined at a workshop held on 27th August 2009 with representatives from each sector and other appropriate organisations. A list of invitees to the workshop can be found in Appendix 8. The output of the workshops was a conceptual model for each of the four engagement types.

The conceptual model for sport is shown in *Figure 40* overleaf. This diagram demonstrates the policy levers and other drivers of doing sport, the outcomes generated by doing sport, and how these factors interact.

The literature and stakeholder engagement also identified a number of cohorts between which the dynamics illustrated in *Figure 40* would vary. For instance, someone of working age is likely to have less free time than others. The complexity of the modelling meant that the number of different cohorts for which separate models were specified were limited to the combinations of following groups:

- Gender: male and female
- Age: 11-15 years old, 16-29 years old, 30-49 years old, 50-64 years old, and over 65 years old.
- Income of the highest earner in the household: low (£0 - £14,999), average (£15,000 - £39,999), high (£40,000+).

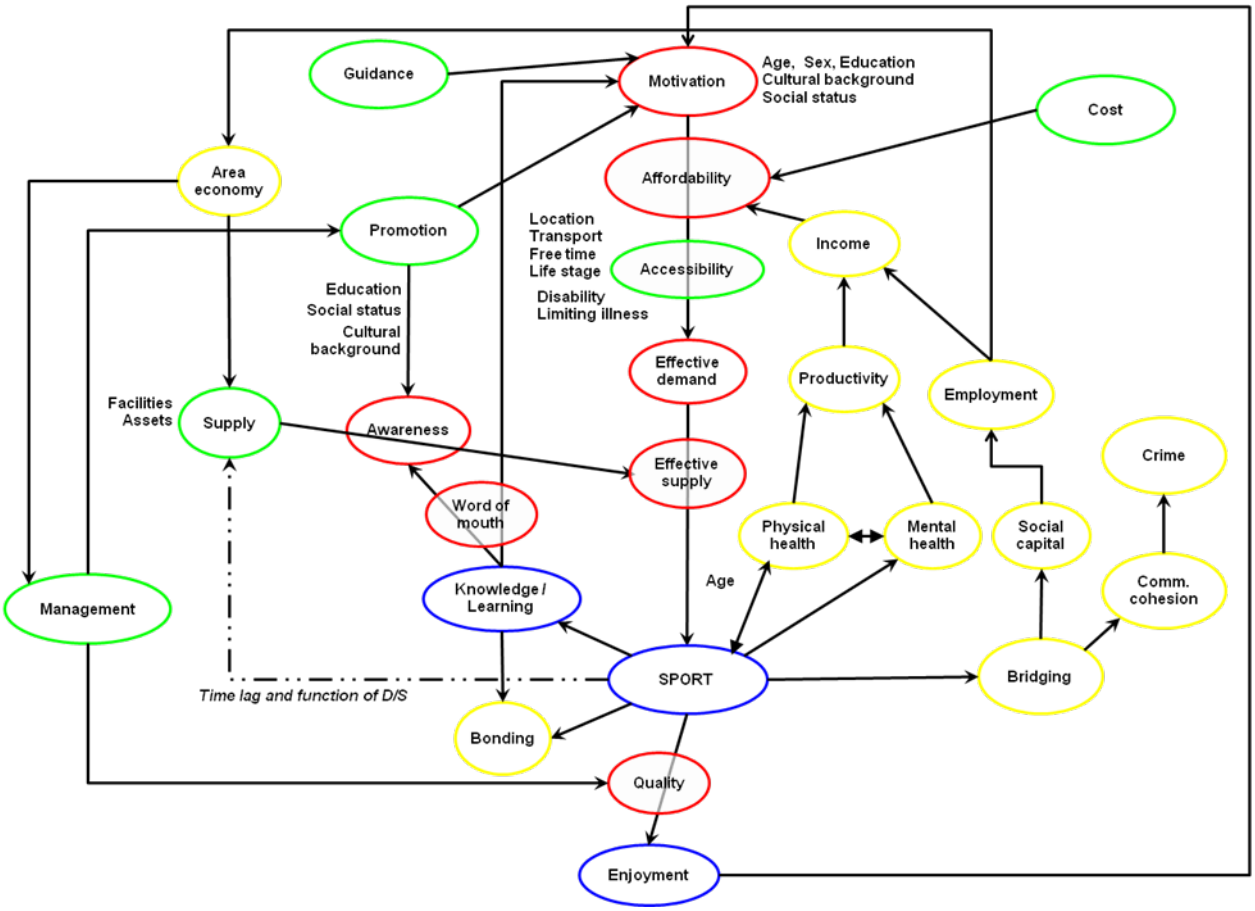
The cohorts included in the model were prioritised based on the following criteria:

- Targets of policies (e.g. certain policies are aimed at children, rather than adults)
- Important demographic trends
- Key factors that influence engagement
- The availability of data (see below)

Table 13: List of activity types modelled

Attendance at arts events	Heritage	MLA	Sport
1. Music	1. A city or town with historic character	1. Museums or galleries	1. Swimming
2. Theatre (adults only)	2. A historic park, garden or landscape open to the public	2. Libraries	2. Health, fitness, gym, conditioning & weightlifting
3. Opera or musical theatre (adults only)	3. A monument such as a castle, fort or ruin	3. Archives	3. Football
4. Opera or musical theatre and theatre (children only)	4. A historic building open to the public (non-religious)	4. All MLA	4. Badminton
5. Visual art	5. A historic place of worship attended as a visitor		5. Golf
6. Street art	6. A place connected with history or historic transport system		6. Athletics (includes track and field athletics, and jogging)
7. Carnival (adults only)	7. A site of archaeological interest		7. Tennis
8. Culturally specific festival (adults only)	8. A site connected with sports heritage		8. Squash
9. Carnival and culturally specific festival (children only)	9. All heritage listed above		9. Cricket
10. Dance			10. Recreational walking
11. Video or digital art			11. Cycling
12. Crafts			12. All sports listed above
13. Books or writing			
14. All arts listed above			

Figure 40: Conceptual model of the factors that influence engagement in sport and the outcomes produced by engagement



KEY

Policy levers

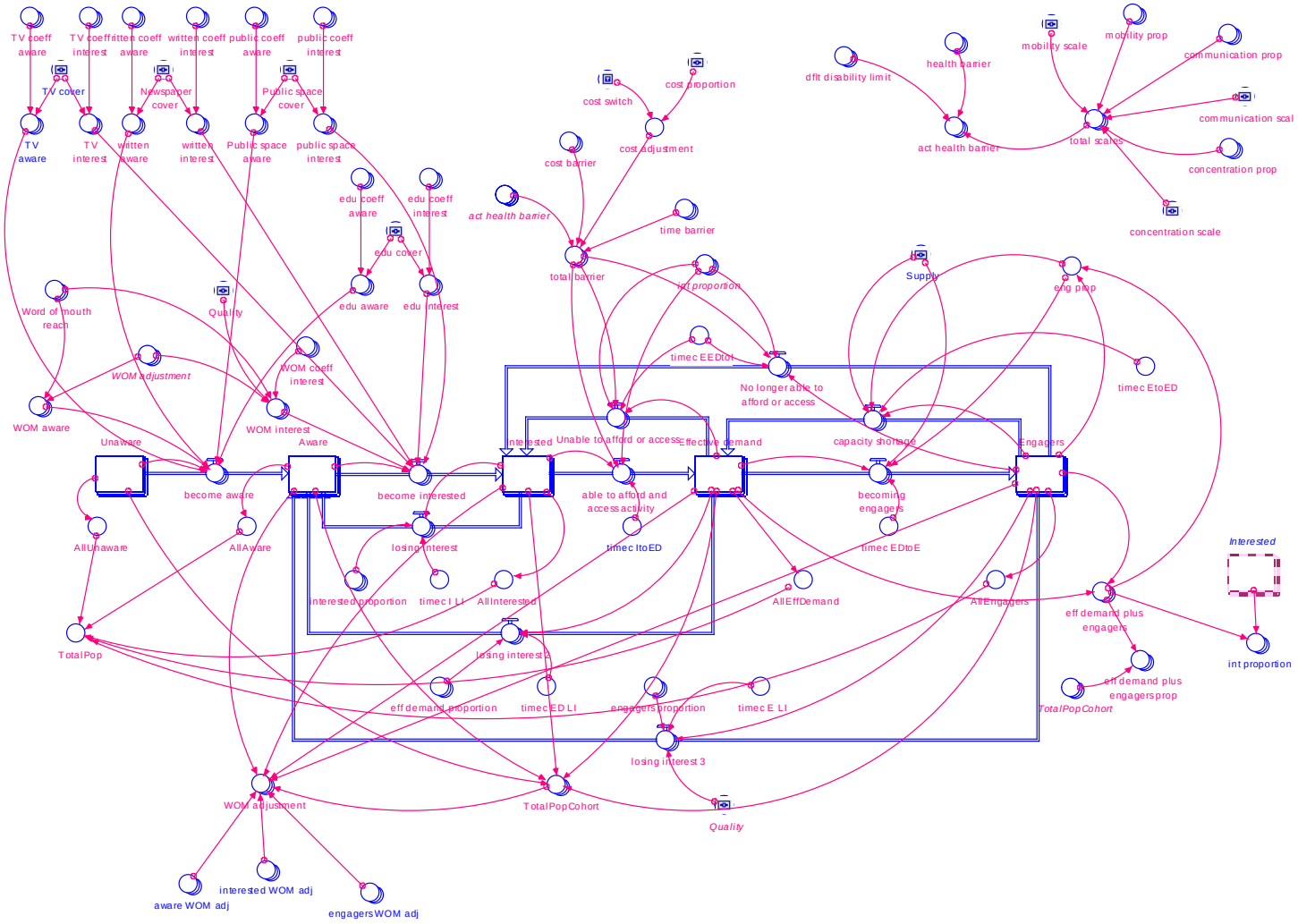
Other drivers

Outcomes generated by all engagement types

Sport specific outcomes

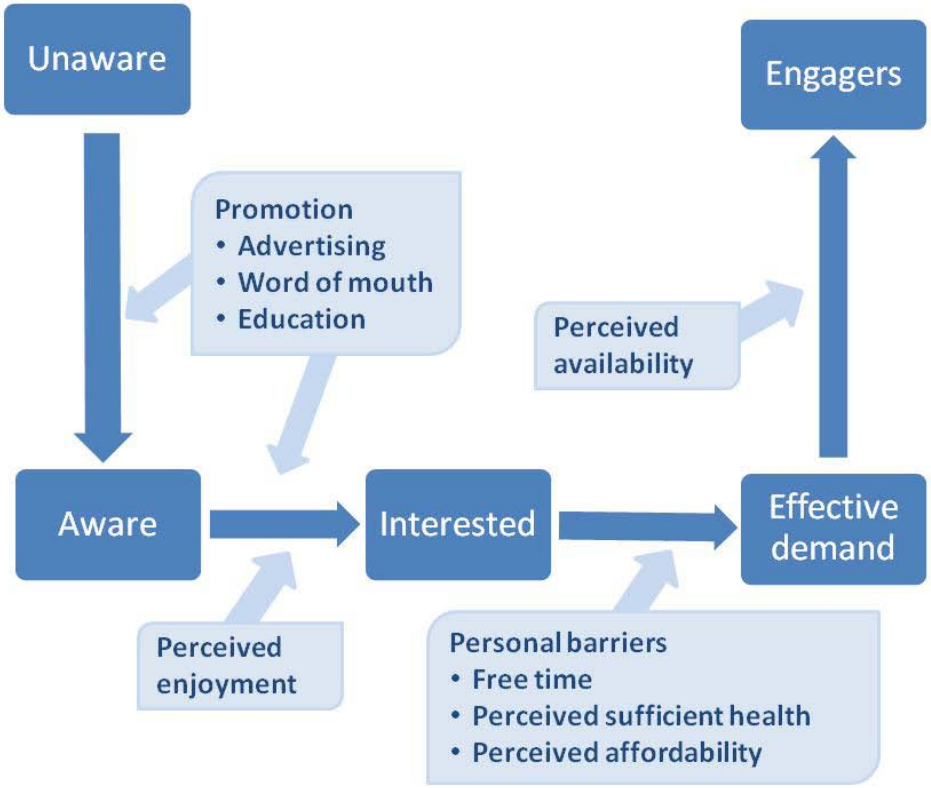
The initial structure of the system dynamics model was specified based on conceptual models. The model was built using software called *ithink* which is designed specifically for system dynamics work. The full model structure from the *ithink* software is shown in *Figure 41* overleaf. This demonstrates how the conceptual model has been converted into the stocks and flows that make up a system dynamics model.

Figure 41: System dynamic model structure from *ithink*



Although the structure in *Figure 41* is not easy to follow, it has been included to demonstrate the complex and dynamic nature of the model structure. *Figure 42* summarises the model structure in a more transparent format. This demonstrates the five key stocks that form the core part of the model, as well as the main drivers that influence whether people flow from one stock to the next.

Figure 42: High-level model structure



At any one time people will sit in one of the following five independent stocks:

- Unaware: People who are not aware that they could engage in the activity.
- Aware: People who are aware that they could engage in a given activity but do not have an interest in engage in the activity.
- Interested: People who would like to engage but are prevented from doing so because either: they can't afford the activity, they don't perceive they have the time, or they suffer from an illness that stops them from doing so.
- Effective demand: People who would like to engage but are prevented from doing so because of a lack of supply or capacity of opportunities.
- Engager: People who have actively engaged in the given activity.

Within the model, individuals are modelled through each of the stocks based on the following drivers:

- Moving from "unaware" to "aware": The model includes the effects of the following factors on peoples' awareness of the opportunities to engage in culture and sport:

  - Promotion: The input to the model is the proportion of the population exposed to the idea of engaging in culture or sport as a result of policies to promote engagement in culture and sport. This could be achieved, for instance, by promotional campaigns via television, the written media, the internet, or public space advertising. The number of people who recall the message and thus remain aware of the opportunity to engage is estimated based on McGuire's "Rule of halves" which is commonly applied in advertising theory (McGuire, 1984). This states that half the people exposed to an advert will recall the message delivered in the advert. The model considers the proportion of people who would be exposed to the campaign. This proportion would vary across different cohorts and would be specific to the given media campaign, for example how widely a billboard poster is displayed.
  - Educational campaign: The input to the model is the proportion of the population exposed to the idea of engaging in culture or sport as a result of an educational campaign. The model assumes that everyone exposed will become aware that they could engage in the activity. Education in respect to children will be more likely to resemble schooling, whereas in adults this could be outreach programmes.
  - Word of mouth: The proportion of people who hear about the opportunity to engage through word of mouth from others in their cohort is estimated as a function of the proportion of current engagers who say that they will tell others about the activity and the size of people's networks of friends. The proportion of engagers who tell their networks about the activity is, in turn, a function of the proportion of engagers who enjoy the activity. These data are derived from the Taking Part survey.

- Moving from "aware" to "interested": The flow between these two states is again a function of exposure to promotional campaigns, educational campaigns, and word of mouth. However, these factors have a different level of influence on becoming interested than they do on being aware. For instance, McGuire's "Rule of halves" states that only 6.25% of people exposed to a message will change their attitude (McGuire, 1984).

- Moving from "interested" to "effective demand" is a function of a number of barriers to engagement. To estimate these barriers, the following inputs are included in the model:
  - The proportion of people who do not have the free time to engage.
  - The proportion of people whose health is not sufficient to allow them to engage. The following five types of limiting illness are distinguished between:
    - Mobility: The proportion of people for whom their mobility prevents them engaging (relevant for all engagement types).
    - Dexterity: The proportion of people for whom their dexterity prevents them engaging (relevant for only the sport models).
    - Physical coordination: The proportion of people for whom their physical o-ordination prevents them engaging (relevant for only the sport models).

        - Communication: The proportion of people for whom communication problems prevent them engaging (relevant for all engagement types except sport).
        - Concentration / memory: The proportion of people for whom problems with their concentration / memory prevent them engaging (relevant for all engagement types except sport).
    - The proportion of people who cannot afford to engage.

- Moving from "effective demand" to "engager" is a function of the supply of opportunities to engage. The input into the model is the proportion of people who do not engage because there is inadequate supply of opportunities.

The model also includes a number of decay functions – flows between stocks in the opposite direction to those discussed above. These were built into the model to take account of the fact that over time individuals change their preferences and behaviours. The following three decay functions are included in the model:

- "Engagers" moving to "effective demand" due to changes in the perceived supply of opportunities
- "Engagers" losing interest and moving to "aware".
- "Engagers"; or those in "effective demand" experiencing a change in personal circumstances and moving to "interested" due to no longer having the time to engage, no longer being able to afford to engage, or no longer being healthy enough to engage.
- Those in "interested" losing interest and moving to "aware".

Data on decay was not available so the decay rates were estimates through the calibration of the model (see section 5.2.4)

Model dynamics

The system dynamics model was run over intervals of one week or a period of 1000 weeks. That is, estimates of the number of people in each stock (unaware, aware, interested, effective demand, and engaging) are recalculated each week. This allows the timing of the effects of policies to be observed. Such timings have not been discussed in this section. However, they will be available for those researchers and policy makers who use the model.

4.2.3 Data

Specifying the model for 38 separate activity types (see table 13 above) for 33 different cohort groups meant that 1,254 separate models were specified. While the structure of the models was consistent within activity types (sport etc), the parameters with the models were estimated for specific activity-cohort groups. The models were constructed to estimate engagement for the population of England.

A number of sources were used to identify data to populate the model, including:

- Analysis of the 2007/8 Taking Part survey
- Reviews of studies provided by members of the CASE programme board[11]
- Studies identified in the CASE database of research (a database of studies on the drivers and value of culture and sport constructed as part of this study[12])
- Supplementary web searches

The details of the data sources reviewed can be found in Appendix 9 and 10.

Owing to the large number of data required to populate the system dynamics model, and the relative paucity of evidence available in the culture and sport fields, it was not possible to populate all model parameters from data identified through the above process. Whilst, for instance, the EPPI database contains many references relevant to policy making in culture and sport, the precise nature of the parameters in the system dynamics model meant that few sources were identified for these parameters. For more detail on the specific model parameters, see Appendix 9.

Where there were gaps in the model, parameters were, for instance, estimated by inferring across cohort/activity types, e.g. the likelihood that an "engager" would recommend an activity was not known for all activities. An estimate of the level of enjoyment was available for all activities. Thus it was assumed that the relationship between enjoyment and the likelihood of recommending an activity was constant across activity levels.

Where data was available to estimate model parameters, much of this data was derived from the Taking Part survey. This emphasises the importance of the survey to the culture and sport fields. However, it also indicates the limitations of the broader evidence-base available in these fields, as well as meaning that the model is strongly exposed to the limitations of the Taking Part survey.

Details of the final data used in the model along with an assessment of the quality of the data can be found in Appendix 9.

System dynamics models are generally considered to be more reliable when it is possible to populate the stocks accurately. The fact that this was achieved using Taking Part data provides confidence in the modelling.

4.2.4 Validating the model

Any model as complex as that reported in this section is subject to range parameter uncertainty. Three standard procedures were used to respond to this uncertainty. First, the quality of the data employed in the model was assessed (see Appendix 9). Second, the impact of uncertainty was tested using sensitivity analysis (see section 5.3.4). Third, the following steps were undertaken to validate the model:

[11] The CASE programme steering committee comprises representatives of the Arts Council, DCMS, English Heritage, MLA (Museums, Libraries, and Archives), and Sport England

[12] Visit the CASE website to use the database: www.culture.gov.uk/case

- Baseline calibration: the level of engagement predicted by the model for each activity-cohort combination was calibrated against the engagement rates estimated from an analysis of the 2007/8 Taking Part survey.
- Change calibration: A literature review was undertaken to identify estimates of the effect of changes in model parameters on engagement levels.
- Sense checking and expert workshop.
- Sensitivity analysis.

In addition, a number of policy scenarios were run through the model – see the case studies reported below. These case studies allow the output from the model to be sense checked, and in some cases – in particular, the scenarios presented for theatres and swimming – they allow the output from the model to be validated against actual evaluation data.

Baseline variance: Once each model had been populated with the available data, it was run without introducing any policy change[13] to estimate the baseline engagement level predicted by the model. This predicted engagement level was then compared against that measured by the 2008/9 Taking Part survey. A key gap in the evidence base required to populate the model was data on decay rates – the rate at which people stop, for instance, being aware of engagement opportunities or being interested in engagement. Thus, decay rates were adjusted through a process of trial and error to calibrate the predicted and actual engagement rates. Appendix 12 summarises the final variance between predicted and actual engagement rates.

Change calibration: Studies of the drivers of engagement in culture and sport – both regression analysis of large datasets and effect studies – were reviewed for evidence on the effect of changing model parameters on engagement levels. Appendix 13 lists the studies included in the review. Unfortunately, this data proved of limited value to the calibration exercise. First, much of the evidence was of limited validity. Second, the studies estimated relationships that were not included in the model. For instance, a number of studies estimated the effect of entrance charges on the number of museum visits. The model, however, does not include a relationship between price and the number of visits. Instead, it includes the relationship between perceived affordability and number of visitors.

Sense checking and expert workshop: Each of the models was run to sense-check that the impact on engagement of changes in model parameters was “reasonable”. In addition, workshops were held with representatives of each of the CASE board members. These workshops involved discussion of the structure and functioning of the models, as well as sense checking of the behaviour of the models.

4.2.5 Using the system dynamics model

The system dynamics model can be used to assess the effects on engagement levels of the following different types of policy:

[13] The baseline policy scenario comprise the following: an absence of promotional campaigns and educational campaigns; and the remaining policy-relevant variables (assets/facility supply and quality, affordability, and disabled access) set as per the existing data. It is likely that this description of the baseline is inaccurate to the extent that, at the time the 2008/9 Taking Part Survey was undertaken, promotional and educational campaigns were in place. It was not possible to measure the existence of such campaigns.

- Promotions, such as communicating information via the internet, or advertising campaigns. Users of the model are required to assess the proportion of the population that the promotion reaches. The model then estimates the effect of this promotion on the numbers of engagers by adjusting proportion of the population who flow between the 'unaware' and 'aware' stock, and between the 'aware' and 'interested' stocks.
- Improved accessibility of facilities and sites, such as providing disabled access. Users of the model are required to assess the impact of a policy on the proportion of the population who do not engage for health reasons. The model then estimates the effect of the policy on the numbers of engagers by adjusting the proportion of the population who flow between the 'interested' and the 'effective demand' stocks.
- Changes in the affordability, such as reduced cost of accessing a site. Users of the model are required to assess the impact of a policy on the proportion of the population who do not engage because they cannot afford to do so. The model then estimates the effect of the policy on the numbers of engagers by adjusting the proportion of the population who flow between the 'interested' and the 'effective demand' stocks.
- Change in the supply of facilities and sites. Users of the model are required to assess the impact of a policy on the proportion of the population who do not engage because of perceived lack of availability of opportunities. The model then estimates the effect of the policy on the numbers of engagers by adjusting the proportion of the population who flow between the 'effective demand' and the 'engagers' stocks.
- Improved quality of experience. Users of the model are required to assess the impact of a policy on the proportion of the population who are satisfied following engaging in the activity. The model then estimates the impact of greater satisfaction on the proportion of population who are recommended they engage in the activity, and on the numbers of engagers by adjusting the proportion of the population who flow between the 'unaware' and 'aware' stock, and between the 'aware' and 'interested' stocks.

The results section provides illustrations of how the model can be used to assess policy outcomes.

4.3 Results

4.3.1 Baseline stocks

Figures 43 to 47 show the baseline proportions of the population in different stocks. That is, before any policies are introduced (see footnote 6), the distribution of the English population across the stocks: unaware, aware, interested, effective demand, and engagers (see section 5.2.2 for definitions of these stocks). These stocks provide useful information for policy makers prioritising investment to improve engagement in culture and sport. For instance, there is little point investing in promotional campaigns if most of the population is already aware and interested in engaging.

Figures 43 and 44 demonstrate a similar "stock-profile" for attending an art event and visiting a museum. Small proportions of the population do not engage because they are either unaware the possibility of engaging or because they perceive supply to be limited. Though in each case a reasonably sized minority sits in these categories and could be impacted on by promotional campaigns.

The majority of non-engagers in both art events and museum visits do not engage because they are uninterested, they don't have the necessary free time, their health limits them doing so, or because they can't afford it.

Figure 43: Population stocks for attendance of an art event

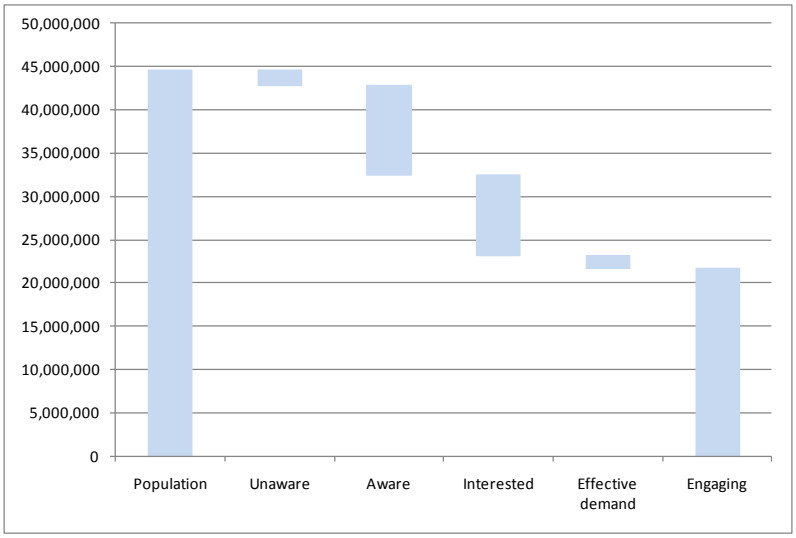


Figure 44: Population stocks for visiting a museum

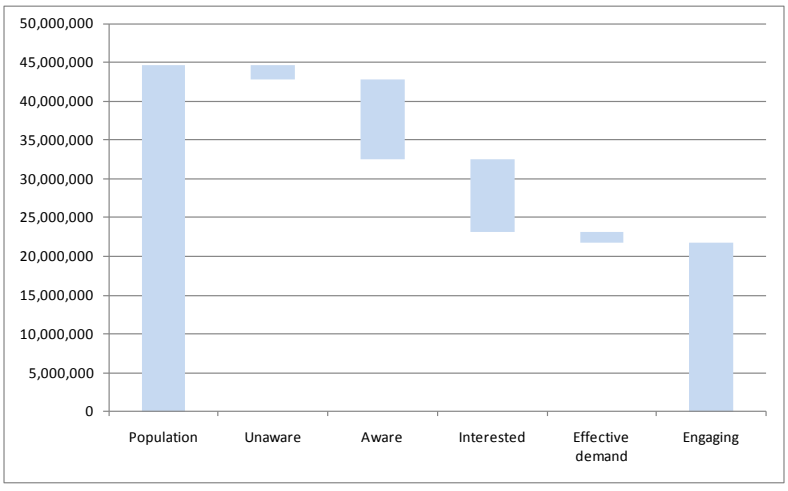


Figure 45 demonstrates a slightly different "stock-profile" for library visits than for museum visit or arts events. In this instance, the majority of non-engagers do not engage because they are not interested. Furthermore, a very small minority of non-engagers do not engage because they perceive the supply of libraries to be limited.

Figure 45: Population stocks for visiting a library

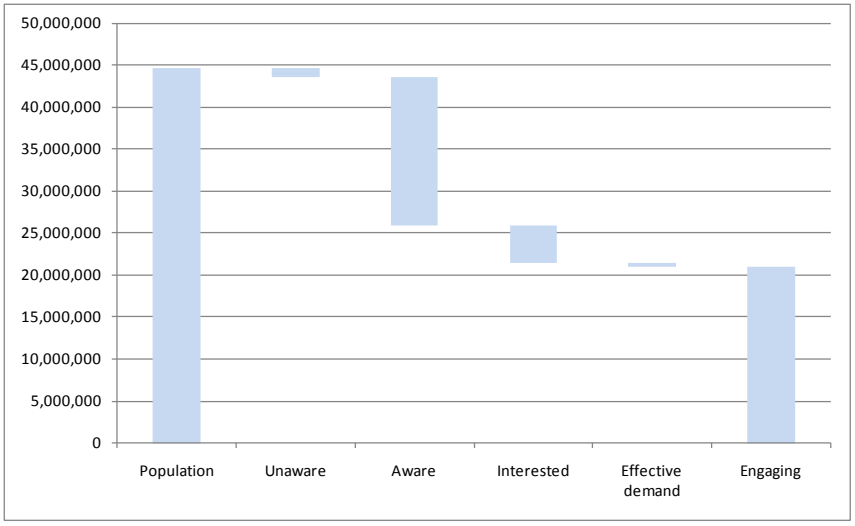


Figure 46 illustrates the "stock-profile" for doing sport. As expected, a very small minority of the population is not aware of the possibility of doing sport. Furthermore, only a very small minority of the population do not do sport because they perceive the supply of facilities/ assets to be limited. Interestingly from a policy perspective, the majority of non-engagers do not engage because they face barriers such as insufficient free time, health limits, or because they can't afford it.

Figure 46: Population stocks for doing sport

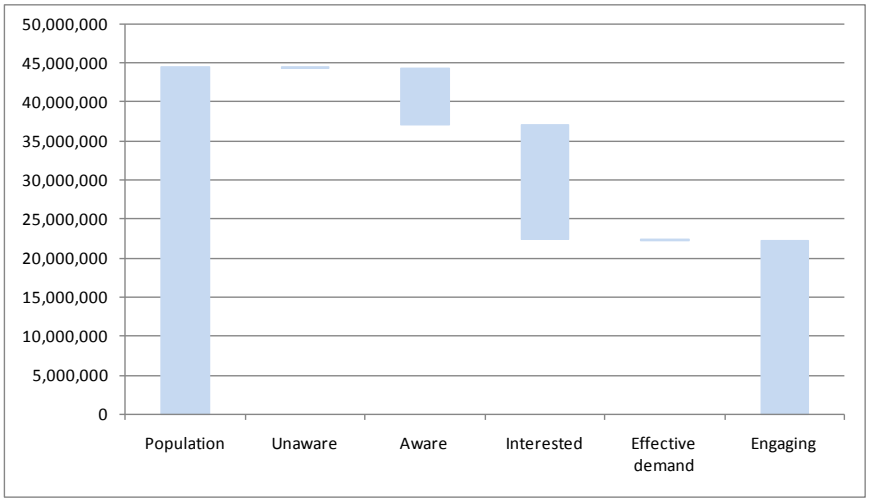
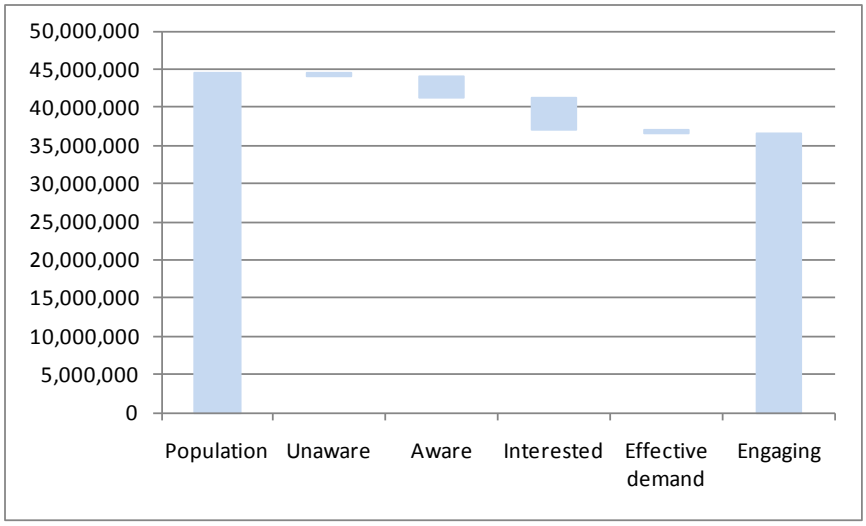


Figure 47 illustrates the "stock-profile" for visiting a heritage site. It demonstrates that a large proportion of the population visit a heritage site in each year. A very small minority of the population don't engage because they aren't aware of the possibility or because they perceive the supply of opportunities to be limited (perhaps because they don't have access to a motor vehicle – see the regression analysis in the previous section). Of the non-engagers, the majority are interested in visiting heritage sites but are prevented from doing so by personal barriers, such as limited health.

Figure 47: Population stocks for visiting a heritage site



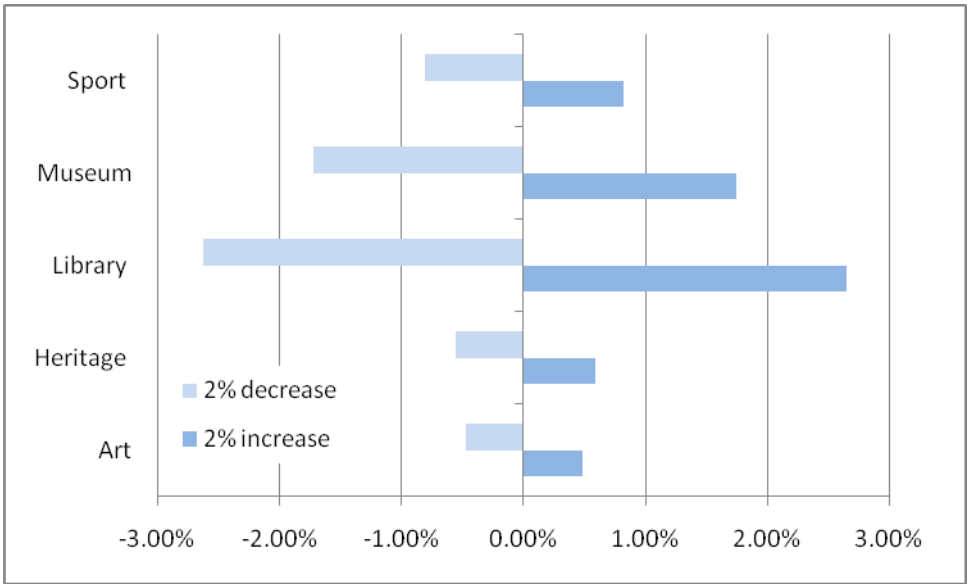
4.3.2 Predicted effect of policy

Figures 48 to 52 show change in engagement level predicted by the model for different policy scenarios. The scenarios are described more fully in *Table 12*. Appendix 14 shows how the changes in engagement levels are distributed across different cohorts.

Figure 48 shows the estimated change in engagement levels with changes in people's level of enjoyment in engaging. More precisely, it demonstrates how engagement is estimated to vary if the assessment of the quality of the experience of engaging increases or decreases by 2%. Such changes could be associated, for example, with policies that invest in the refurbishment of sports facilities or the information provided at heritage sites, museums or galleries. Within the model, this effect is produced by affecting the likelihood that engagers recommend the activity to their networks.

The direction of the changes in *Figure 48* is as expected – improvements in the quality of experience of engaging results in a higher level of engagement, and vice versa. *Figure 48* also demonstrates the relative sensitivity of engagement to quality of experience across the different sectors. Specifically, visits to libraries and museums are more sensitive to the quality of the experience than engagement in the other culture and sports sectors. This finding is consistent with the results of the regression analysis reported in the previous section. That analysis found that only engagement in libraries was associated with the perceived quality of facilities. Furthermore, the sensitivity of engagement to perceived quality is associated with the proportion of the population aware but interested in engagement. For libraries, this group make up about 40% of the population. It is reasonable to speculate that improving the quality of experience of engagement will increase the chance that engagers recommend the activity to their networks, increasing the probability that this group moves from being just aware to being interested.

Figure 48: The impact of improved quality of experience on engagement levels (percentage change relative to baseline)[14]



[14] In the model, improved quality of experience operates through the effect on word of mouth – increasing the likelihood that engagers will recommend an activity to their friends, increasing the chance that these people will become aware and/or interested.

Table 14: Description of policy scenarios

Policy scenario	Before	After	Description
Improved quality of experience	79%	81%	A policy is introduced to raise the quality of experience so that the % of users who state they enjoyed the activity increases by 2%
Reduced quality of experience	79%	77%	The quality of experience is allowed to reduce so that the % of users who state they enjoyed the activity decreases by 2%
Improved access for those with limited communication skills	Varied default*	100%	A policy is introduced so that all those with limited communication skills can access the activity
Improved access for those with limited concentration/memory	Varied default*	100%	A policy is introduced so that all those with limited concentration/memory can access the activity
Improved access for those with limited physical coordination	Varied default*	100%	A policy is introduced so that all those with limited physical co-ordination can access the activity
Those not able to afford the activity is set at 0%	Varied default*	0%	A policy is introduced so that everyone can afford to participate in the activity
Those not able to afford the activity is set at 10%	Varied default*	10%	A policy is introduced so that only 10% cannot afford to participate in the activity
A promotional campaign reaches 10% of the population	0%	10%	A promotional campaign is introduced that reaches 10% of the population
Reduce supply/capacity	-	-10%	The supply/capacity of assets/facilities is reduced so that 10% of the population perceive there to be insufficient supply to allow them to participate when they did not think this previously

* The baseline setting within each activity – cohort model will vary

Figure 49 shows the estimated effect on engagement levels of removing the barriers to engagement for those with limiting illnesses. The physical disability barriers limit the number moving from being interested in engaging to moving into the effective demand stock. Improving the access allows a greater number of people to move through into the effective demand stock. Once in this category, they will be subject to the same perceived supply barriers as those already in this stock. Some will therefore move through into active engagers, whilst others will remain in the effective demand stock.

Reducing barriers to those with limited mobility is estimated to have a substantial positive impact on engagement in culture and sport. The most dramatic effect of a policy to remove barriers to those with limited mobility is estimated to be on participation in sport, which would increase by 15%. Removing the barriers for those with limited physical co-ordination, and dexterity is also estimated to have a significant impact on participation sport, which is estimated to increase by 5%.

Given the way in which the model is constructed, inevitably the effect of reducing barriers is correlated with the proportion of those who state they are interested in an activity type but that health limits their ability to engage in the activity. For instance, 60% of those who are interested in sport state that health limits their ability to engage. This compares with 13%-20% for the other activity types. Furthermore, it is not a surprise that reducing barriers to engagement for those with limited mobility is the most effective strategy identified in *Figure 49*. Of those people who report a health barrier as limiting their ability to engage in an activity, the majority of these people (60%-73%) report limited mobility as a limiting factor.

Figure 49: The impact of reducing barriers to engagement on engagement levels (percentage change relative to baseline)

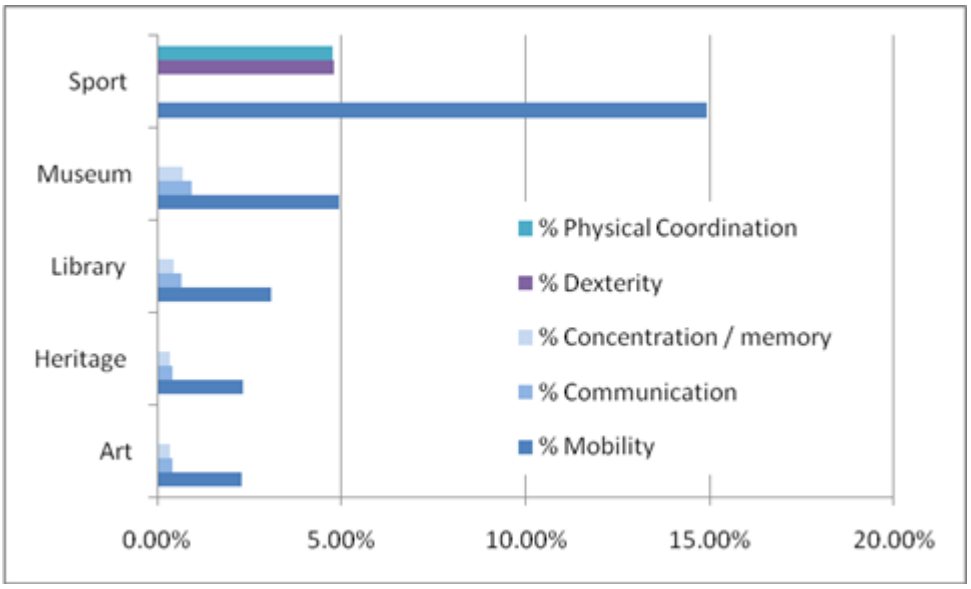


Figure 50 shows the estimated effect of policies targeted at the perceived affordability of engagement (where affordability refers to the total cost of engaging, including entrance cost, travel and equipment). It shows the effect of two scenarios: a policy that makes engagement perceived to be affordable to all (and a policy that results in 10% of the population still unable to afford the activity. It demonstrates that ensuring that everyone can afford to engage is estimated to have a small positive effect on engagement levels across the sectors. The greatest estimated effect is observed on the numbers of people doing sport and visiting museums. A policy or other change in context such as a recession that resulted in 10% of the population not being able to afford to engage in culture and/or sport would reduce engagement levels.

As with the disability barriers, changes to the perceived affordability impacts on the number of people that move between the interested and effective demand stocks. In the case of everyone being able to afford to engage, as people have move into effective demand, a proportion of these will also move to become engagers, whilst for others, perceptions of availability will prevent them moving further and they remain in effective demand.

Figure 50: The impact of perceived affordability on engagement levels (percentage change relative to baseline)

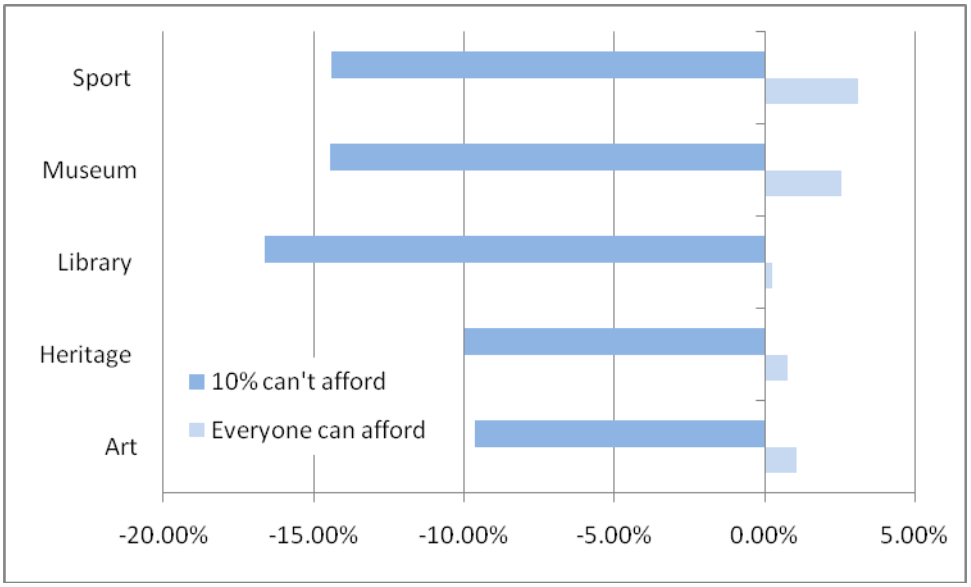


Figure 51 shows the estimated impact of promotional campaigns on engagement levels. Specifically, it shows the effect of a promotional campaign that is seen by 10% of the population. It predicts that such a campaign would have a varied effect on engagement across the sectors. Promotional campaigns impact on the model at a variety of points, impacting not only on the awareness but also on the level of interest in participation. Only very small effects are observed for visiting heritage sites and attending arts events. In these sectors, there are a greater number unaware than in the other sectors. The advertising campaign moves some of these people into awareness, but these do not all immediately progress to being interested in participating. By contrast, a large effect is observed for visits to libraries, with the number of people visiting increasing by 17%. This might be partly due to the larger number of people who are aware of libraries where the advertising has the effect of moving these individuals to being interested (see *Figure 49*). The individuals moving into

the interested categories are subject to the same barriers as those already in these stocks, such that some of those individuals who were interested now move through to become engagers over the modelling period.

Figure 51: The impact of promotion on engagement levels (percentage change relative to baseline)

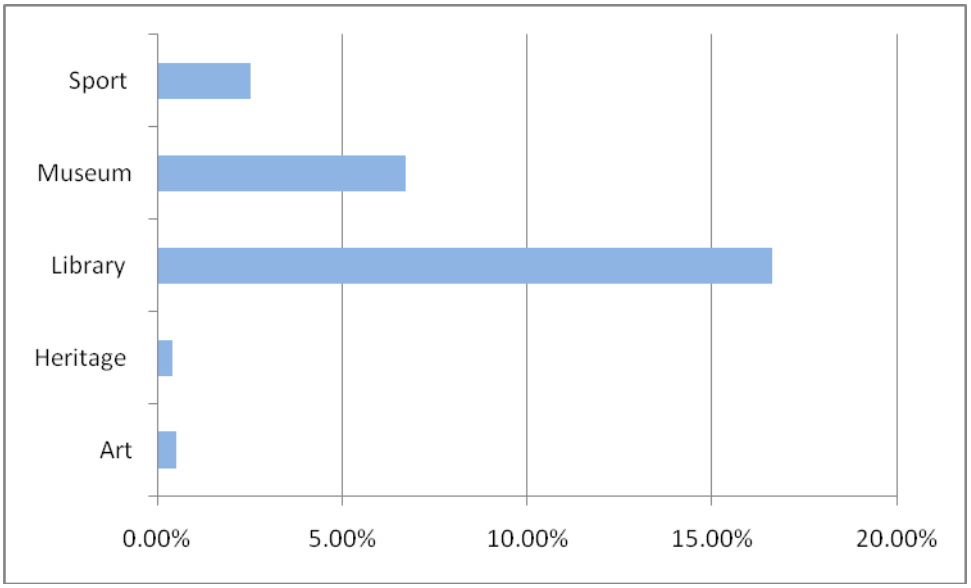
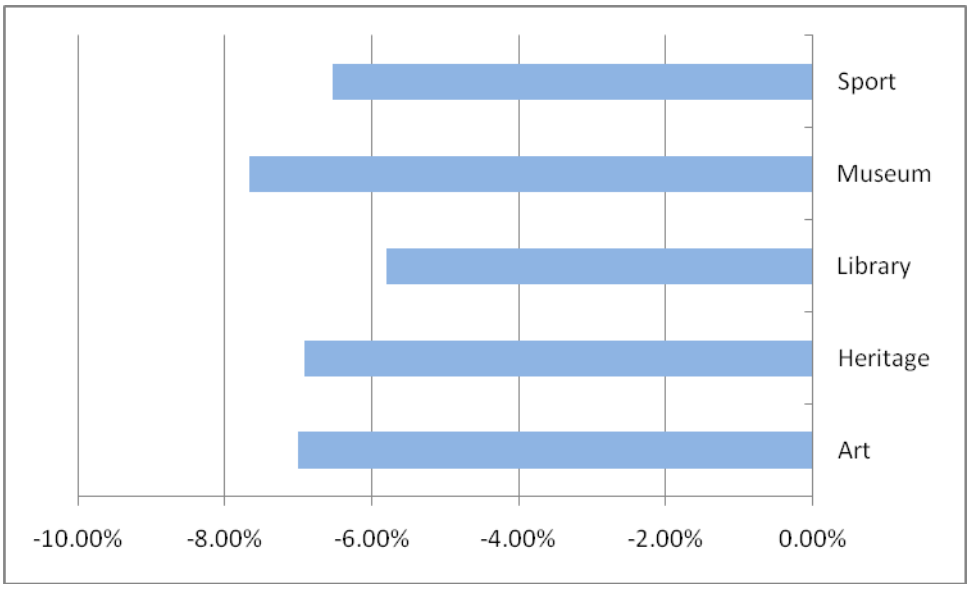


Figure 52 shows the estimated effect of a reduction in the supply/capacity of assets/facilities on engagement levels. Specifically, it shows the effect of an increase of 10% in the proportion of the population who think there is insufficient capacity to allow them to participate in the activity. The effect of this change is to move individuals who had previously engaged to move from being engagers to being in the effective demand stock. A similar effect is estimated across the sectors, with engagement levels dropping approximately 6%.

Figure 52: The impact of reduced capacity/supply on engagement levels (percentage change relative to baseline)



Case studies 1-5 summarise the results of modelling more specific sector-relevant policy scenarios. These not only illustrate how the model can be used to inform policy making, but also act as a real work sense check on the outcome of the models. Appraisal of the likely impact of policy using the model requires information on how the policy impacts on key model parameters. For instance, appraisal of a policy to change the cost of museums entry requires information on how this policy impacts on the perceived affordability of accessing museums. This information will need to be collected through primary evaluation work. For further information on using the model, see the Users' Manual that accompanies this report.

Case study 1: Policy to improve attendance at art events

Policy: The provision of reduced-price theatre tickets for people aged 16 to 24.

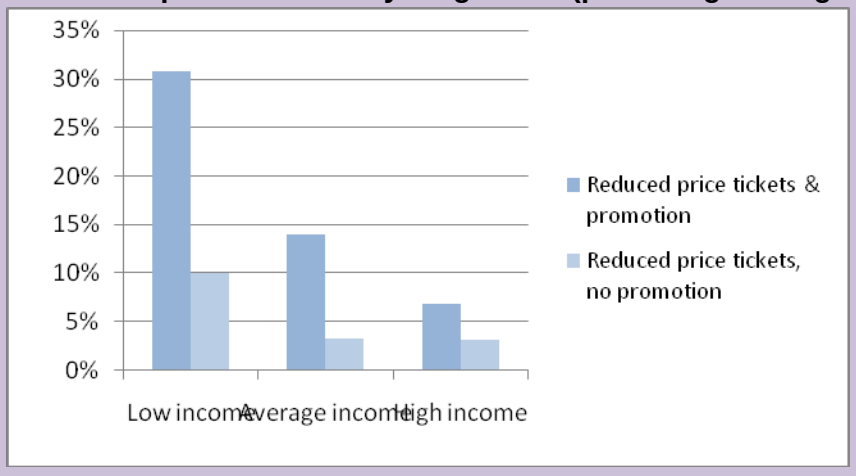
Modelling approach: As the reduced-price tickets are available only to a limited cohort, the model was run for the cohort aged 16–30. The policy impact was modelled through changing the perceived affordability of attending theatre for individuals in this cohort. In the baseline model, perceived affordability ranged between 3% and 7.5% for this cohort. In order to model reduced price tickets the perceived affordability barriers were reduced to half of the baseline levels for the associated age groups. Two scenarios were run:

- Perceived affordability set to half of the baseline level
- Perceived affordability set to half of the baseline level with a promotional campaign that reaches 5% of the population.

Two scenarios were run: the introduction of the reduced ticket price intervention with and without promotion.

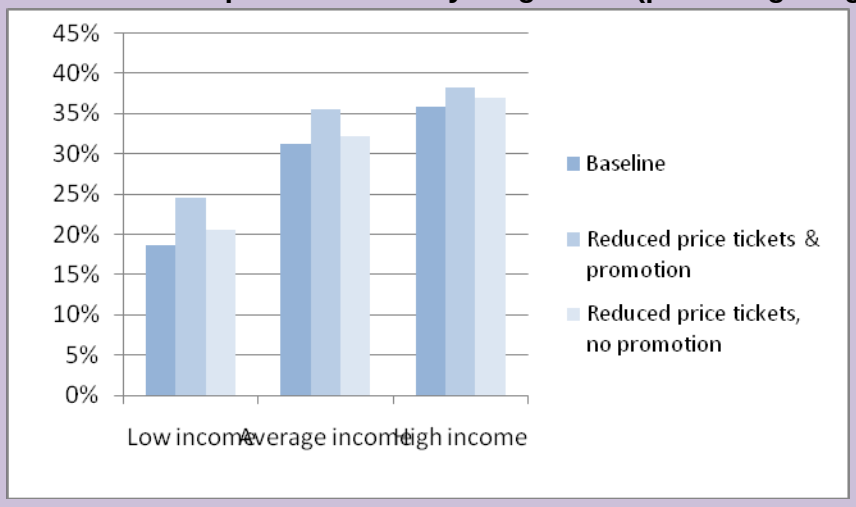
Result: The figures below report the results of the modelling. Unsurprisingly, it is estimated that the effect of the reduced theatre ticket price policy is greatest when combined with promotion. Equally unsurprisingly, it is estimated that the effect of the policy is greatest on the attendance of those in low-income groups.

The effect of reduced price tickets on young males (percentage change from baseline)



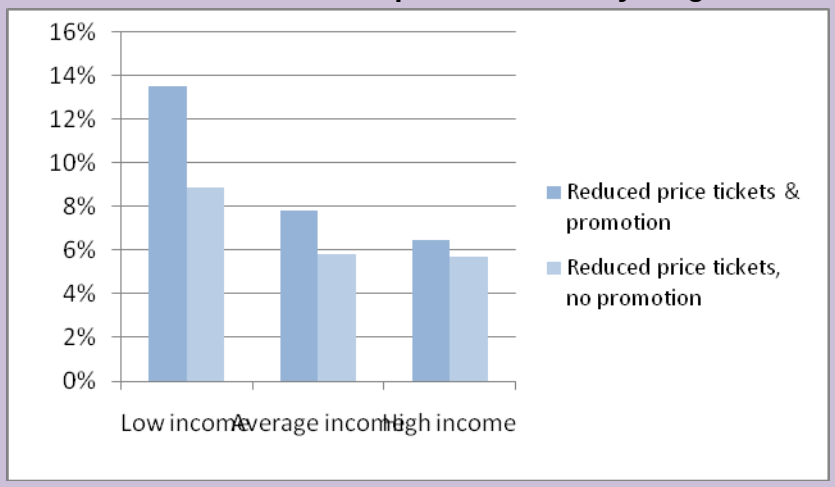
For instance, it was estimated that the combined price reduction and promotion policy would increase the number of young men on low incomes attending the theatre by about 30%. This compares with a 10% increase in attendance amongst this group for free theatre without promotion.

The effect of reduced price tickets on young males (percentage engagement)

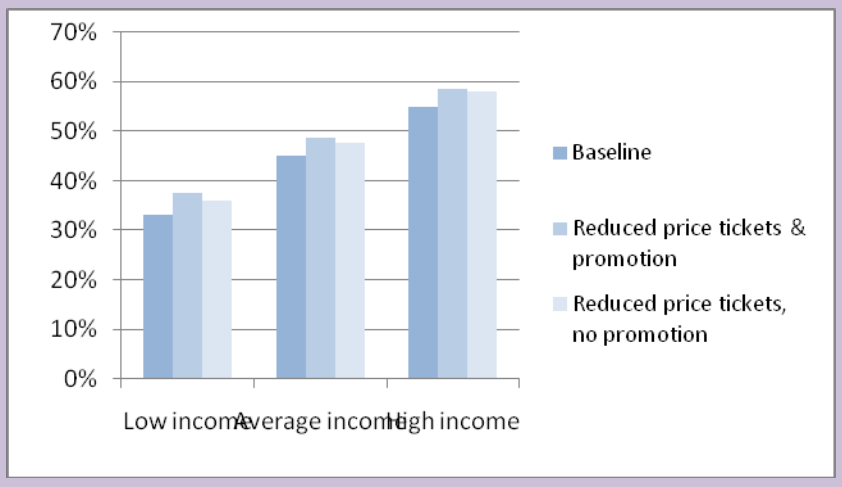


The effect of the policy was estimated to be lower for females than for males. It was estimated that the combined price reduction and promotion policy would increase the number of young women attending the theatre by about 14%.

The effect of reduced price tickets on young females



The effect of reduced price tickets on young females (percentage engagement)



Case study 2: Policy to improve visits to heritage sites

Policy: Campaigns to increase the number of visitors to historic buildings. Two campaigns are compared with one another: an educational campaign that reaches 10% of the population; and a promotional campaign that reaches 10% of the population.

Modelling approach: The historic building model was run twice. In each case, the model was run for all cohorts.

Results: It was estimated that the increase in the number of people visiting historic buildings would be 4.9% following the educational campaign and 4.3% following the promotional campaign. This percentage difference in the increase in the number of visitors to historic buildings between educational and promotional campaigns held across gender and income groups.

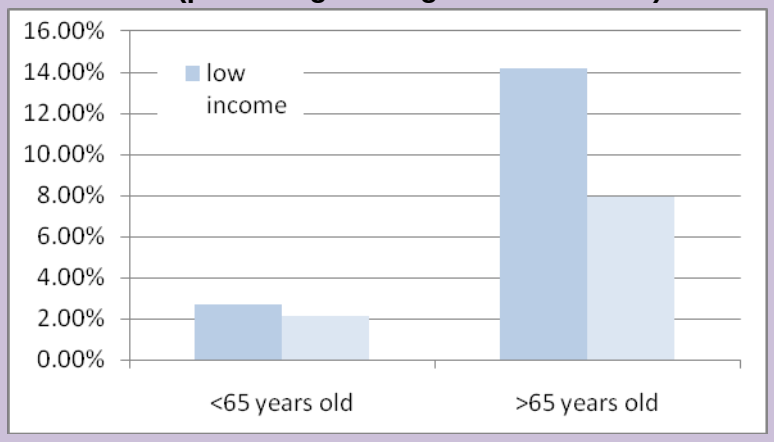
Case study 3: Policy to improve visits to libraries

Policy: Mobile libraries to encourage greater use of library facilities. This is particularly beneficial for cohorts who have problems accessing facilities. The mobile libraries might also result in an increase in the opening hours of libraries, and will reduce the time to access a library as individuals do not have to travel so far. It could also help those whose health is a barrier to allowing them to access facilities.

Modelling approach: The model was run for the whole population, with a 5% reduction in

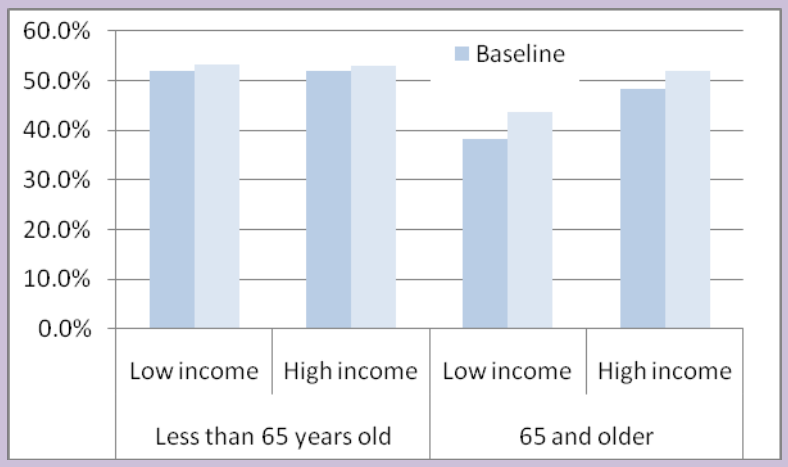
the percentage of each cohort for whom time was a barrier to visiting a library. It was apparent that, while for younger people health was generally not a barrier to visiting libraries, this was not the case for the older populations. Therefore, for the over-65 cohort, a 5% reduction in the percentage of people reporting health as a barrier to visiting a library was also applied.

The effect of mobile libraries on the number of people making use of libraries (percentage change from baseline)



Result: It was estimated that the intervention would increase the number of people making use of libraries by 3.8%. The effect on visitor numbers varied between age groups and income levels. Unsurprisingly, the effect on the number of people over 65 years old visiting libraries was greater than for those under 65 years old. The effect on the number of people in lower income groups visiting libraries was greater than on those in higher income groups.

The effect of mobile libraries on the number of people making use of libraries (percentage engagement)



Case study 4: Policy to increase visits to museums

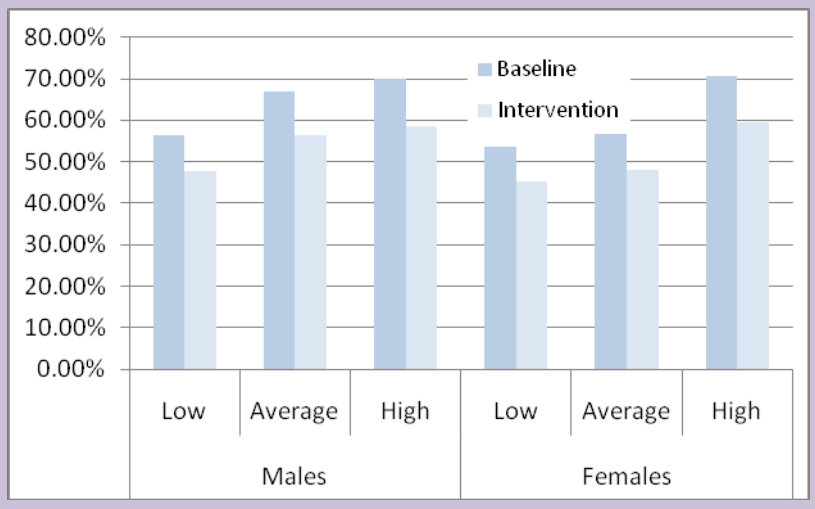
Policy: A 20% reduction in the provision of funding to museum or gallery related education.

Modelling approach: This policy would affect children attending school, so the model was run for the cohort aged 11-15 years old. The effect of the policy was modelled in by estimating the difference in visitor numbers for two scenarios: the introduction of a standard educational promotion; and one that is slightly reduced in effectiveness (to simulate a reduction in funding). Two sets of scenarios were run, both reducing the reach of education by 20%: a reduction in reach from 10% to 8%; and a reduction in reach from 25% to 20%. The scenarios produced the same percentage change in visitor numbers.

Result: As expected, the effect of a reduction in education funding is a reduction in the number of young visiting museums, with a 20% reduction in the reach of educational campaigns producing a 15-16% reduction in the number of visits among young people.

The effect of reduced education funding on young people visiting museums (percentage change from baseline)

The effect of reduced education funding on young people visiting museums (percentage engagement)



Case study 5: Policy to improve participation in sport

Policy: The DCMS is currently running the free swimming initiative in local authorities across England. The initiative is aimed at young people aged 16 years or under and older people aged 60 years and over. As the initiative is not being run across all local authority areas and the free swims are generally offered only for specific sessions, affordability issues haven't been completely removed but they are greatly diminished for these groups.

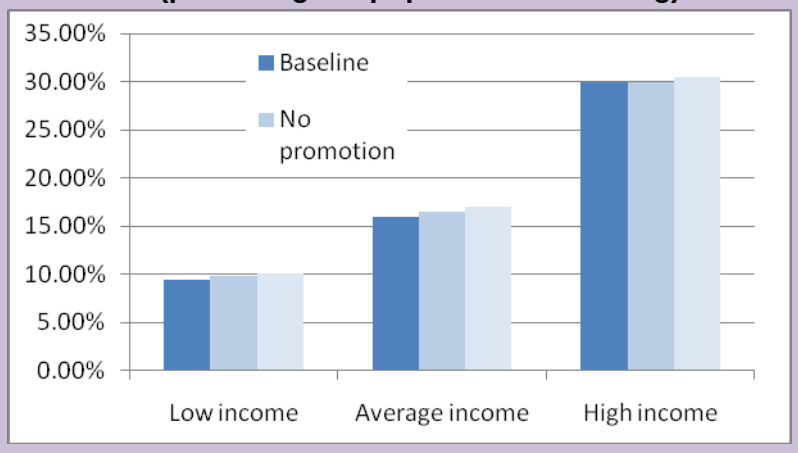
Modelling approach: The models for the 11-15 years old and the over-65 years cohorts were run. Although not all local authorities will implement free swimming, the model does not operate at a local geographical level. The national-level model was run on the assumption that the local authorities where the intervention is being rolled out are representative of the nation as a whole, and thus the results of the national-level model are applicable. In order to model free swimming the perceived affordability barriers were reduced to half of the baseline levels for the associated age groups. Two scenarios were run:

- Perceived affordability barriers were set to half of the baseline level
- Perceived affordability barriers were set to half of the baseline level with a promotional campaign that reaches 5% of the population.

Affordability barriers were not removed as entrance costs are only one component of the cost of swimming. Other costs include, for instance, travel costs.

Results: It is estimated that the free swimming initiative will increase the total number of people who swim in these age groups by 0.5% without promotion and 0.8% with promotion. Proportionately, the largest impact is found in the low income cohorts with a 4.7% and 7.3% (no promotion and 5% promotion respectively) increase on the baseline engagement levels. In terms of the actual increase in swimmers, the largest impact is seen in the average income cohorts with a 25,000 and 48,000 increase in the number of swimmers.

The effect of free swimming and promotion on the number of people who swim (percentage of population swimming)



The effect of free swimming and promotion on the number of people who swim (change in actual number of swimmers)

4.3.3 Predicted effect of socio-economic trends

This section presents the estimated effects on engagement levels of the following two socio-economic trends:

1. Population projections for 2012 (ONS, 2008): the baseline population profile of the models was changed to reflect changes in the proportions of people in different gender and age groups.
2. Gross Domestic Product (GDP) projections for 2012[15]: the baseline population profile in the model was changed to reflect changes in the proportion of people in different income groups.

Appendix 15 summarises the results of modelling these scenarios in more detail.

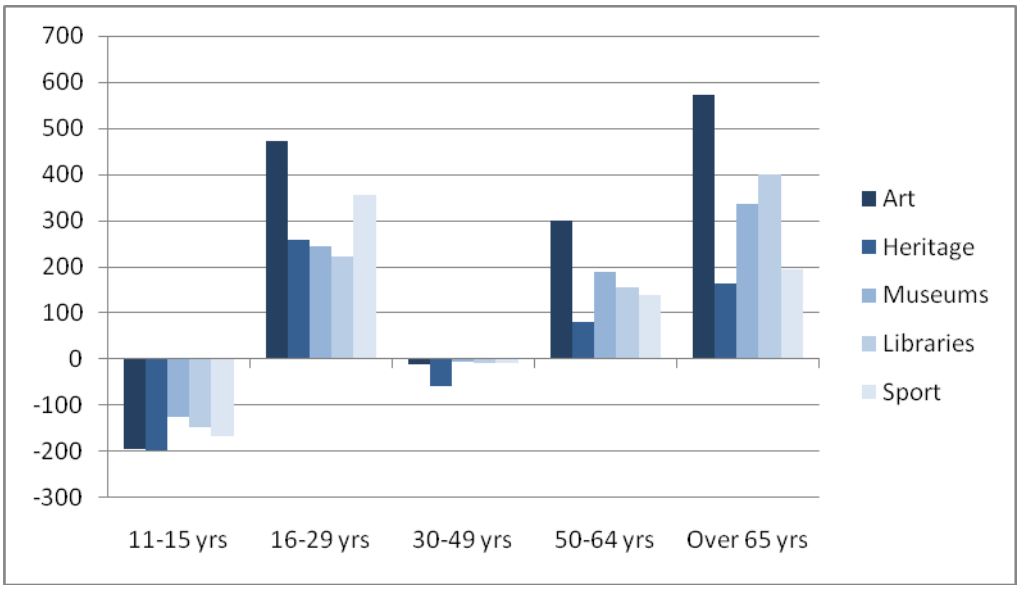
The effect of changes to the age profile of the population

[15] H.M. Treasury projections for private consumption (http://hm-treasury.gov.uk/d/pbr09_annexa.pdf)

It is estimated that the effect of changes in the age of the population by 2012 will increase the numbers of people attending arts events, visiting museums and visiting libraries by about 3%, increase the numbers of people doing sport by 2.3%, and increase the number of people visiting heritage sites by 0.7%.

Demographic scenarios are implemented in the model by moving people into different age cohorts based on the demographic projections. The model assumes that people entering the cohort are distributed between stocks in the same way as those already in the cohort, and that they exhibit the same behaviour as those already in the cohort. The change in the engagement levels are driven by the different baseline engagement levels in each cohort. *Figure 53* shows the effect of changes in the age profile of the population on the numbers of people engaging in culture and sport within different age cohorts.

Figure 53: Change in the number of people engaging in culture and sport as a result of demographic changes by 2012 (thousands)



In addition to modelling changes in population to 2012, analysis was also carried out to model the effect of population change up until the year 2020. It is estimated that the population changes predicted over this period will cause an increase in engagement across all age groups across all sectors. However, this is most notable in attendance at arts events in the 50 – 64 and over 65 years age groups.

The effect of GDP changes

It is estimated that the effect of changes in GDP by 2012 will increase the numbers of people engaging in culture and sport, but only by small amounts – between 0.05% (heritage and libraries) and 0.15% (museums). GDP scenarios are implemented in the model by moving people into different income cohorts based on new income figures. The model assumes that people entering the cohort are distributed between stocks in the same way as those already in the cohort, and that they exhibit the same behaviour as those already in the cohort. As the GDP changes result in an increased number of people in the higher income groups (where

engagement is higher) there is an overall increase in the number of engagers. That is, the change in the engagement levels are driven by the different baseline engagement levels in each cohort. *Figures 54a* and *b* show the effect on the numbers engaging in each of the sectors.

Figure 54a: Change in the number of people engaging in culture and sport as a result of demographic changes by 2020

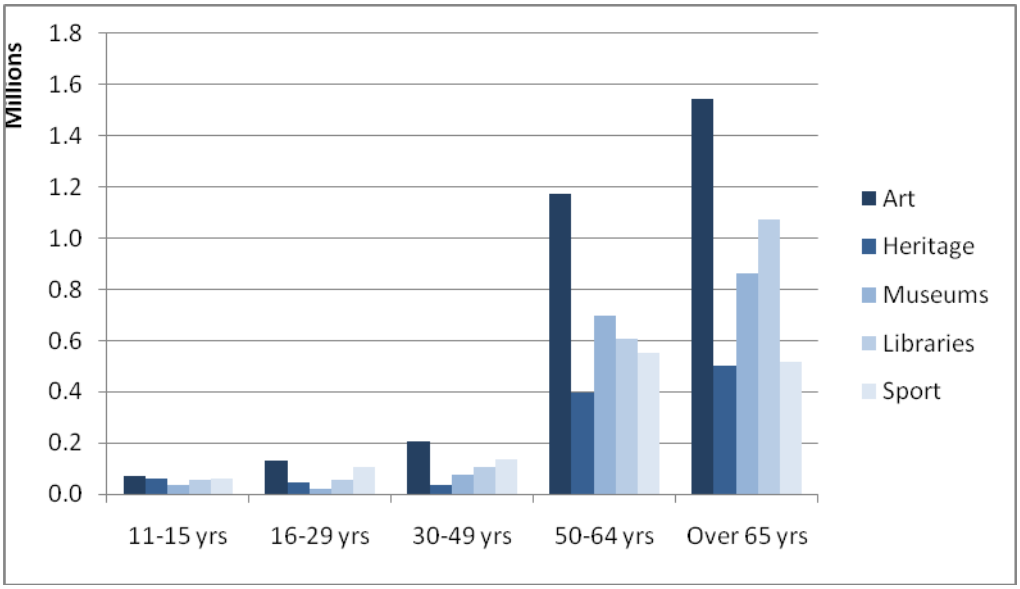
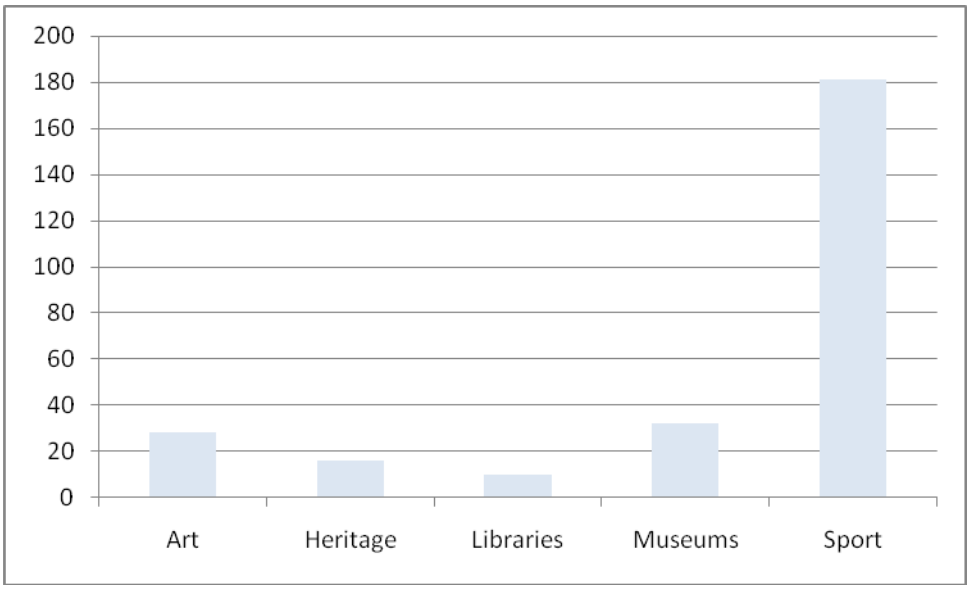


Figure 54b: Change in the number of people engaging in culture and sport as a result of GDP changes by 2012 (thousands)



4.3.4 Sensitivity analysis

Figures 55 to 59 illustrate the results of a one way sensitivity analysis to assess the effect of uncertainties in model parameters on model predictions. The sensitivity analysis was run for the four variables about which there was greatest uncertainty:

- The impact of on whether an individual becomes interested in engaging after having had an activity recommended ("word of mouth interest coefficient").
- The proportion of those interested in engaging who lose interest.
- The proportion of those demanding engagement who lose interest.
- The proportion of engagers who lose interest.

The sensitivity analysis assessed the effect of increasing and decreasing each of these parameters by 5% and 10%.

Each of the figures demonstrates the very small effect of variations in model parameters on model predictions.

Figure 55: Sensitivity analysis – percentage change in the number of people attending arts events

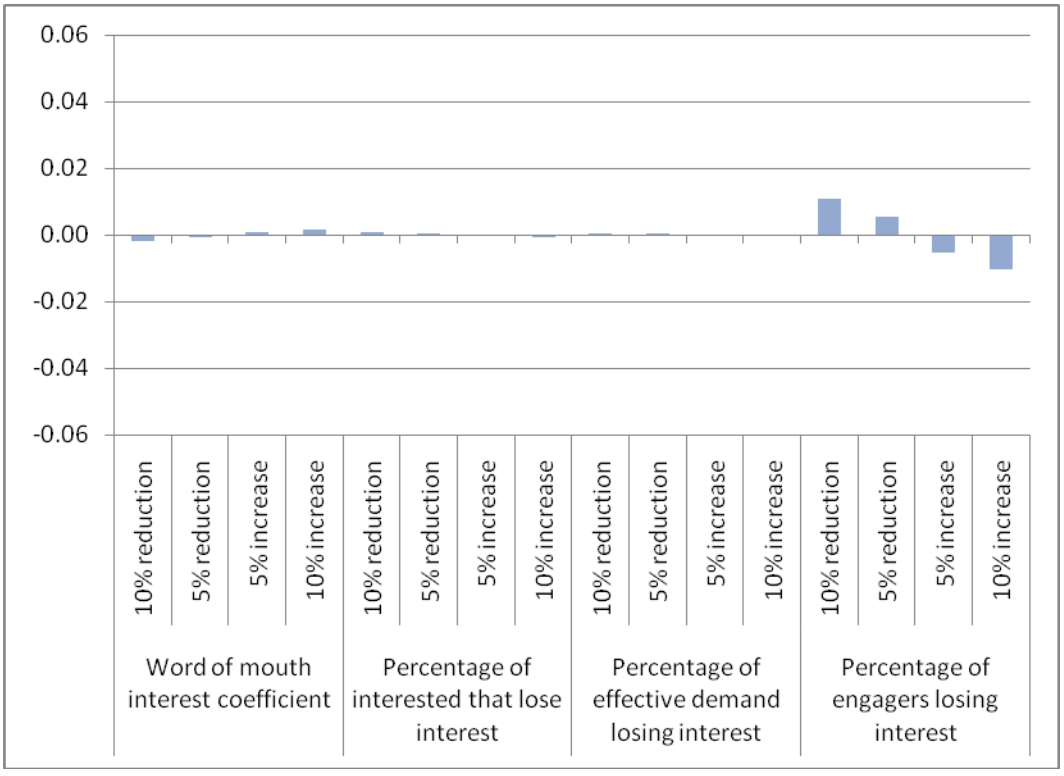


Figure 56: Sensitivity analysis – percentage change in the number of people visiting heritage sites

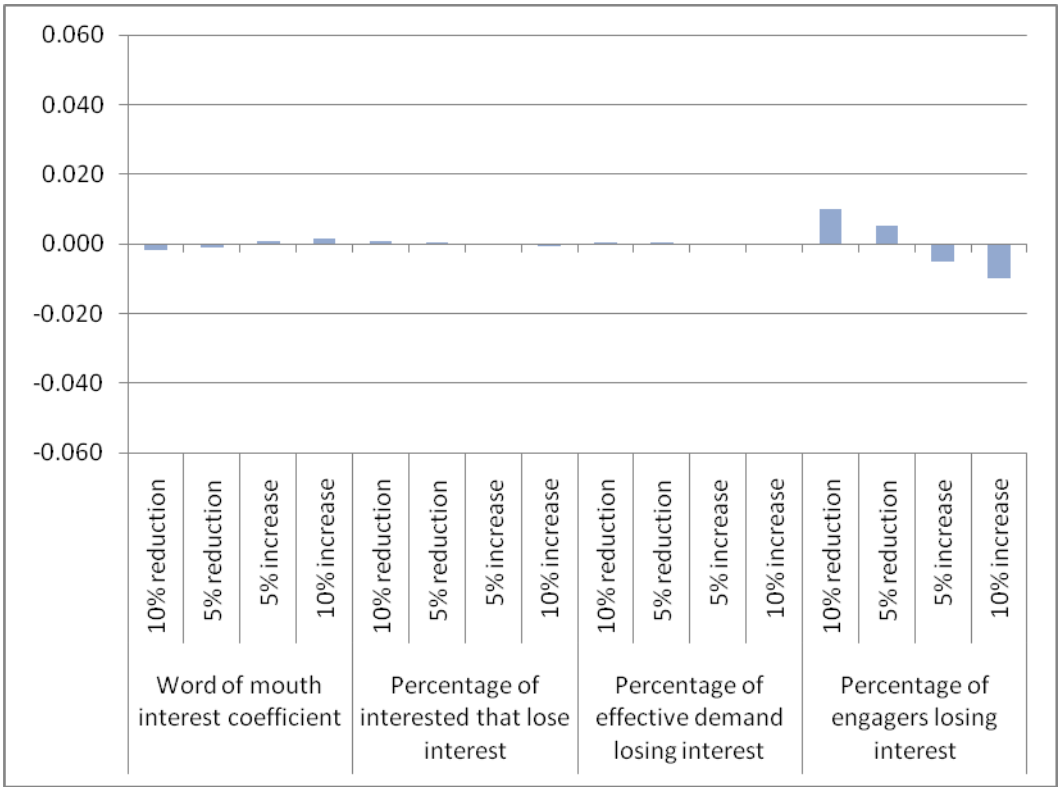


Figure 57: Sensitivity analysis – percentage change in the number of people visiting libraries

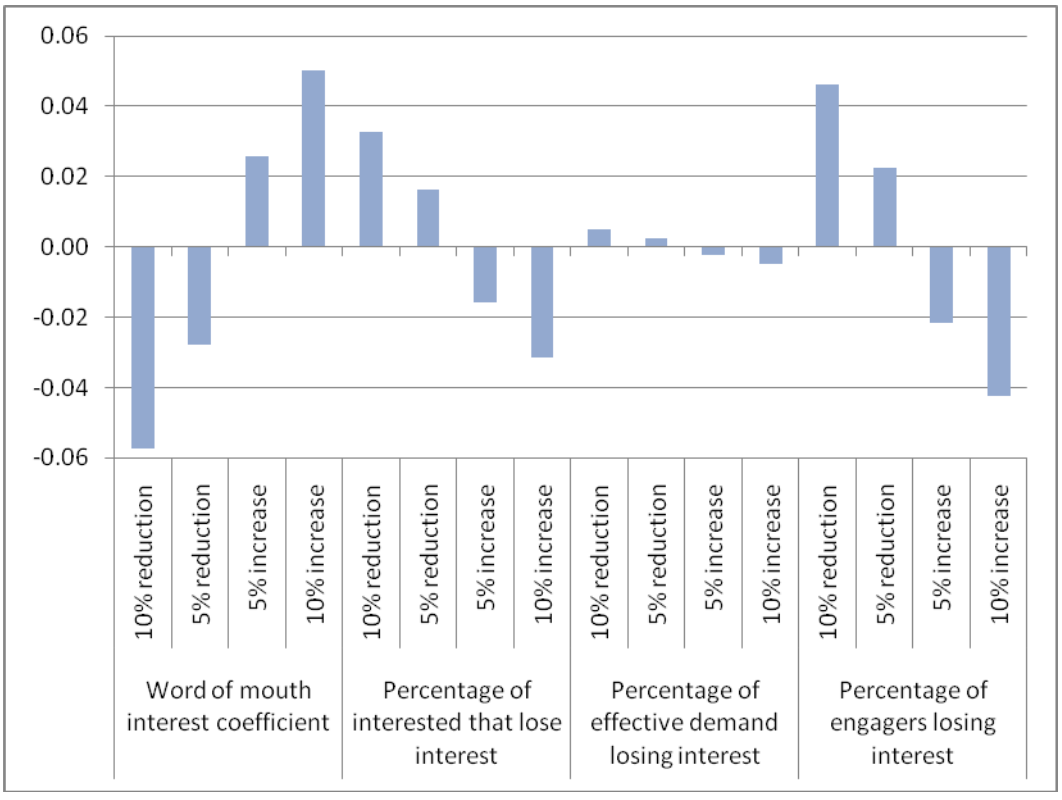


Figure 58: Sensitivity analysis – percentage change in the number of people visiting museums

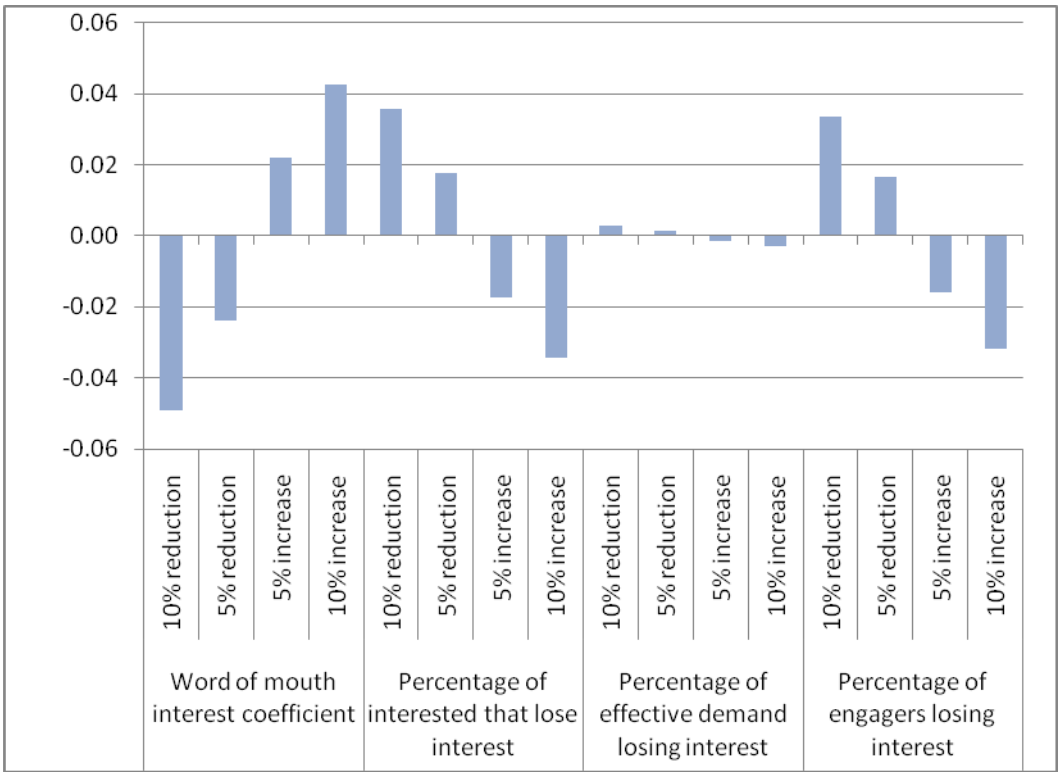
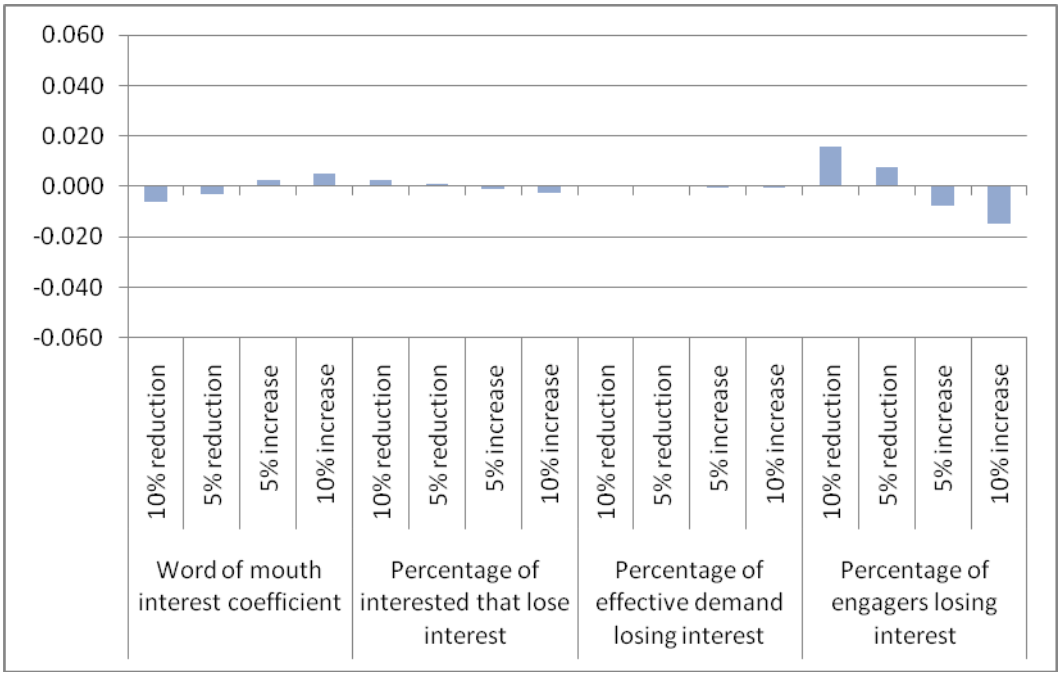


Figure 59: Sensitivity analysis – percentage change in the number of people doing sport



4.4 Discussion

What drives engagement in culture and sport? Section three identified a number of concerns with the conventional regression analysis approaches to answering this question. First, the data available for such analysis is limited in its ability to measure all relevant drivers. Second, the structure of regression models fails to replicate the sequence of decisions necessary before people engage in culture or sport. As a consequence, it is argued that regression analysis fails to provide the insight into the drivers of engagement in culture and sport required to inform policy making. This section summarised a system dynamics model constructed to overcome these challenges. What role do such models have in the development of policies to increase engagement in culture and sport?

Implications for policy making

The models summarised in this section provide a number of insights into the effects that policies can have on engagement in culture and sport. First, modelling people through the sequence of decisions required before they engage in culture and sport provides policy makers with insights into the immediate reasons for non-engagement. A number of interesting observations are available from this exercise, including:

- Only a small proportion (<5%) of people are interested in engaging, face no personal barriers (such as affordability), and don't engage only because of limitations in the supply and/or capacity of existing facilities/assets. By implication, policies aimed solely at increasing the supply and/or capacity of facilities/assets will have only a limited effect on engagement levels.
- With the exception of visiting heritage sites, large proportions (c40-50%) of those who don't engage are either aware but not interested, or interested but for reasons of either time pressure, limiting illness or affordability do not engage. Which of these "states" people inhabit will have important policy implications. For instance, the library and sports sectors have similar numbers of people in these two states, but the distribution of people between the two states varies between the two sectors. A greater proportion of non-engagers in sport are interested in doing sport, but unable to do so for health or financial reasons, or because of competing calls on their time. The opposite is the case for libraries, for which a greater proportion of non-engagers are aware of the opportunities to visit libraries, but are not interested in doing so.

Second, the model estimates that socio-economic trends over the next two years will increase the number of people engaging in culture and sport. Specifically, it is estimated that changes in the age of the population will increase engagement between 0.7% (visits to heritage sites) and 3% (visiting a library); and changes in GDP will increase engagement between 0.05% (visits to heritage sites) and 0.15% (visits to a museum).

Third, the model estimates the effect of a number of policy outcomes on engagement in culture and sport, including:

- Policies to impact people's satisfaction with the experience of engaging will have a greater affect on the numbers engaging in museums and libraries than in other sectors.
- Policies to increase the perceived affordability of engagement will only have small effects on the numbers engaging.
- Promotional campaigns to increase awareness and interest will have a greater effect on the numbers engaging in libraries and museums than in other sectors.
- Policies to reducing barriers to the participation of people with limiting illnesses could potentially have large effects on the numbers of people engaging in culture and sport. This is particularly the case for policies targeting communication problems to increase the number of people doing sport.

The models presented in this section are designed to test the impact of a range of policy outcomes on the engagement levels of a range of population cohorts. The analysis presented in this section represents only an illustration of the outputs that can be generated by the models.

Implications for research

The factors that drive engagement in culture and sport are multiple and complex, their effect varying between different cohorts of the population. The policies designed to increase the numbers engaging in culture and sport vary in their design, the context within which they are implemented, the cohort they target, the cultural/sporting activity they target, and their geographical coverage. As a consequence, predicting the effect of policy on engagement levels will also be a very complex and difficult challenge.

The challenges created by this complexity are faced by policy makers on a day-to-day basis. Constructing a model of such complex policy problems provides the opportunity to break up the problem into its theoretically coherent parts, make these parts explicit and open to discussion, assess the evidence available for each of these parts, and provide guidance about the best solutions to the problem. The model summarised in this section does this for policies to increase engagement in culture and sport. It makes explicit the steps that precede the decision to engage in culture and sport, and in doing so provides a much greater evidence base on which to make decisions.

The paucity of the evidence base in the field of culture and sport meant that populating the model was a challenge. Where data was available to estimate model parameters, much of this data was derived from the Taking Part survey. This emphasises the importance of the survey to the culture and sport fields. However, it also indicates the limitations of the broader evidence-base available in these fields, as well as meaning that the model is strongly exposed to the limitations of the Taking Part survey. Further research is required to develop better estimates in the areas covered by the model. Following best practice, sensitivity analysis was employed to test the effect of parameter uncertainty on model outcomes. This provided some comfort about the validity of the model outcomes, demonstrating the very small effect of variations in model parameters on model predictions.

Despite specifying 1,254 different iterations of the model (for a range of different activities and cohorts), it was necessary to make a number of simplifying assumptions when constructing the model. Each of these assumptions represents an area for further research that could improve the value of the model, including:

- **Engagement types**: Types of engagement in culture and sport were limited to those most prevalent in the English population, as measured by the 2007/8 Taking Part survey. The models include 35 different engagement types. This, however, leaves a large range of engagement types still to model. Development of the model should consider incorporating more engagement types.
- **Cohorts**: Cohorts included in the model are limited to age groups, income groups and gender. However, policy often targets other cohorts, such as families and BME groups. Further work is required to extend the model to allow the impact of policies on the engagement of these other cohorts to be analysed.
- **Effect studies**: The model draws on effect studies for two purposes. First, to validate of the predictions of the model. Second, the use of the model requires estimates of the impact of policies on model parameters, such as increased affordability, improved satisfaction or engagers, and improved access. Reviews of the evidence identified few studies that could be used for these purposes. Further evaluations of the effect of policies to increase engagement in culture and sport are thus required. These should focus on the effect of policies on the different types of stock and policy levers included in the model.
- **Model parameters**: Empirical research is required to improve the quality of the parameter estimates included in the model. This could focus on the variables on which there is least research and those variables that the model results are most sensitive to. Parameters for which least research was identified include the effects of promotional campaigns on awareness and interest, and the rate at which people move backwards down the stocks.
- **Model structure**: The models of each activity-cohort combination were constructed independently of each other. The fact that many of the activities in the models can probably be considered substitutes implies that ideally one model should be constructed that takes into account such interactions. Further model development is required to introduce this extra complexity.
- **Local decision making**: The models are currently constructed at a national-level, estimating the number of engagers for the whole of England. A next step would be to model smaller area levels, such as local authorities, where many culture- and sport-related decisions are taken. The possibility of constructing this model is currently limited by the data available at these area levels. For instance, a key source of data for the model was the Taking Part survey. Once sub-group analysis has been undertaken by, for example, age groups, the Taking Part data is not as robust at a local authority level. Further research to collect culture and sport engagement data at local level will allow the model to be extended to this level.

5. Summary and conclusion

The objective of this research was to answer the question: What drives engagement in culture and sport? The answer to this question has important policy implications as it will directly inform the policy approaches of the Department for Culture, Media and Sport (DCMS) and associated bodies take in broadening engagement. That is, a better understanding of the drivers of engagement in culture and sport will ensure policy is focused on the interventions that are most efficient at increasing engagement.

Section 3 summarised the extant literature on the drivers of engagement in culture and sport. It is argued that the econometric literature is limited by its focus on the conventional economic parameters of price and income. Specifically, the econometric literature overlooks key socio-demographic factors that influence engagement in culture and sport, such as the amount of time required to engage in culture and sport, and the importance of education and previous experience in shaping preferences for culture and sport.

Furthermore, section 3 argued that the approach which predominates in the extant literature (both economic and other science) – a focus on regression analysis of large survey and administrative datasets – suffers from important limitations when analysing the drivers of engagement in culture and sport. First, existing survey and administrative data does not measure a sufficient proportion of the policy-relevant drivers of engagement. Second, the structure of regression models fails to replicate the sequence of decisions necessary before people engage in culture or sport.

Sections 4 and 5 presented empirical research to understand the drivers of engagement in culture and sport. Section 4 summarised the results of a regression analysis similar to that conventionally adopted in the social science literature. Two innovations were applied to address some of the criticisms of this literature. First, a range of datasets were reviewed and analysed to maximise the drivers of engagement which were included in the analysis. Second, sophisticated multi-level modelling techniques were applied to capture both area-level and individual-level effects. Third, the analysis was undertaken in a way that allows comparison across the culture and sport sectors.

The regression analysis confirmed the importance of a number of socio-demographic factors identified in the literature as influencing demand for culture and sport, including education, socio-economic status, and childhood experience of culture and sport. These factors, however, have little relevance for short-term policy objectives to increase engagement in culture and sport. The analysis did, however, include a number of other factors more amenable to policy influence. First, few associations were identified between the quantity and/or quality of cultural and sporting sites and facilities and the probability of people's engagement in culture and sport. This observation is, however, probably the result of challenges in measuring the quantity and quality of such assets.

Second, a negative association was identified between having a limiting illness and the probability of doing sport, visiting a museum and attending an art event. This might imply that improvements the access of these groups to museums and art events, such as

improving transport and making museums and art locations more user friendly, could increase engagement.

Third, a positive association was identified between media access and the probability of engagement in culture and sport. This would suggest that a policy of improved access to media will have positive effects on engagement. Furthermore, to the extent that media access is a proxy for access to information on cultural and sporting opportunities, this might also suggest that promotional campaigns will be effective in increasing engagement.

Section 5 summarised the results of a system dynamics model designed to draw on a wider range of data than that amenable to regression analysis and to reflect the sequence of decisions required before someone engages in culture or sport. The models provided a number of important policy insights:

- Policies aimed at either increasing the supply and/or capacity of facilities/assets or at improving the affordability of engagement will have only a limited effect on engagement levels.
- Policies to increase interest in engagement and to remove barriers to engagement, such as health limitations, are more likely to be successful at increasing engagement levels. The relative effectiveness of these policies will, however, vary between sectors. For instance, the number of people doing sport is likely to be more responsive to policies that remove personal barriers to engagement, while the number of people visiting libraries is likely to be more responsive to policies to promote interest in libraries.
- Policies to impact people's satisfaction with the experience of engaging will have a greater effect on the numbers engaging in museums and libraries than in other sectors.

The system-dynamic models also suggest that socio-economic trends over the next two years, especially changes in the age of the population, will increase the number of people engaging in culture and sport. Specifically, it is estimated that changes in the age of the population will increase engagement between 0.7% (visits to heritage sites) and 3% (visiting a library); and changes in GDP will increase engagement between 0.05% (visits to heritage sites) and 0.15% (visits to a museum).

The innovative approaches summarised above provide important insights with which to inform policy making to increase the number of people engaging in culture and sport. However, these policy areas are relatively poorly served by research and evidence. Therefore, there are a number of avenues of enquiry that would provide a more solid footing for future policy making, including:

- Longitudinal data collection would allow greater insight into the factors that cause people to engage in culture and sport. An alternative would be to incorporate questions on engagement in surveys such as the British Household Panel Survey.
- Theoretical research on the mechanisms by which decisions to engage are arrived at and the role of drivers in influencing these decisions.
- Data collection on the proximity, quality and cost of cultural and sporting opportunities.

- Local area-level data on engagement rates for important sub-groups of the population.
- The effect of policy interventions on key steps in the sequence of decisions required to engage, such as becoming interested in engagement, or overcoming barriers to engagement.
- Evidence on the interaction between engagement types – whether, for instance, engaging in one type of sport and/or culture acts as a substitute or complement for engagement in other sports and/or cultures.

6. References

Aguiar, M., & Hurst, E. (2006). Allocation of time over five years. NBER Working Papers 12082

Anderson, R.G.W. (1998). Is charging economic? *Journal of Cultural Economics*, 22.

Arthur, D. (2008). Diffusion modelling and industry dynamics in mobile telecoms data services. University of Surrey: PhD Thesis.

Becker, G.S. (1965). *A theory of the allocation of time. The Economic Journal* (Vol. 75, No. 299, pp. 493-517).

Bourdieu, Pierre (1984). *Distinction*. London: Routledge and Kegan Paul.

Brook, O., Boyle, P. and Flowerdew, R. Forthcoming (2005). "Geographic analysis of cultural consumption. In Stillwell, J., Norman, P., Thomas, C., and Surridge, P. (eds.), *Understanding Population Trends and Processes. Volume 2: Population, Employment, Health and Well-being*. Springer Publishing.

Bunting et al. (2008). From indifference to enthusiasm: Patterns of arts attendance in England. London: Arts Council.

CEBR – Centre for Economics and Business Research (2007). Main technical report: Attending heritage sites: a quantitative analysis of data from the Taking Part Survey. London: English Heritage.

Chan, T.W., & Goldthorpe, J.H. (2005). The social stratification of theatre, dance and cinema attendance. *Cultural Trends* 14 (3).

Chan, T. W., & Goldthorpe, J. H. (2007a). Social stratification and cultural consumption: music in England. *European Sociological Review* 23 (1).

Chan, T. W., & Goldthorpe, J. H. (2007b). Social stratification and cultural consumption: The visual arts in England. *Poetics* 35.

Christakis, N.A. & Fowler J.H. (2009) Connected: the surprising power of our social networks and how they shape our lives. Little Brown & Co.

Coalter. F (2002), Let's make Scotland more active: entrance charges and sports participation: a review of evidence. Centre for Leisure Research, University of Edinburgh.

Cowell, B. (2007). Measuring the impact of free admission. *Cultural Trends* (16: 3).

David, K. (2004). Sport and early learning experiences. In *Sport England, driving up participation: The challenge for sport*. London: Sport England.

DCMS (2007) *Culture on demand: Ways to engage a broader audience*. London: DCMS.

Frey, B.S., & Meier, S. (2003).The Economics of Museums. In Ginsburgh, V.A. and Throsby, D. (eds.), *Handbook of the economics of art and culture, vol*. *I*. Elsevier B.V.

Gayo-Cal, M., Savage, M., & Warde, A. (2006). A cultural map of the United Kingdom, 2003. *Cultural Trends* (15:2/3, 213–237).

Harris, J.R. (1999) The nuture assumption: why children turn out the way they do. Free press.

Haworth, J.T., & Veal, A.J. (2004). *Work and Leisure*. London: Routledge.

Johnson, P.S. (2003). Museums. In Towse, R. (Ed.), *A handbook of cultural economics*. Cheltenham: Edward Elgar.

Kay, T. (2004). The family factor in sport: A review of family factors affecting sports participation. In *Sport England, driving up participation: The challenge for sport*. London: Sport England.

Kesenne, S. (2006), Public policy and participation sport. In Rodriquez, P., Kesenne, S. and Garcia, J. (eds.), Sports economics after 50 years: Essay in honour of Simon Rottenberg. Universidad de Oviedo

Lera-Lopez, F. and Rapun-Garate, M.A.F. (2007), The demand for sport: Sport consumption and participation models. Journal of Sports Management, 21

Lévy-Garboua, L., & Montmarquette, C. (2003). Demand. In Towse, R. (Ed.), *A handbook of cultural economics*. Cheltenham: Edward Elgar.

MacDonald, M., & Rascher, D. (2000). Does bat day make sense? The effect of promotions on the demand for major league baseball. *Journal or Sport Management* (14).

McCarthy, K. F., et al. (2001). Guide to the literature on participation in the arts. RAND.

MLA (2005). Renaissance Museum development fund audience development grants 2004/05. London: MLA.

MLA (2008). Research briefing 4: Increasing attendance and participation. London: MLA.

Oskala, A., Keaney, E., Chan, T. W., & Bunting, C. (2009). Encourage children today to build audiences for tomorrow: Evidence from the Taking Part survey on how childhood involvement in the arts affects engagement in adulthood. London: Arts Council.

PLB Consulting Limited (2001). Developing new audiences for heritage. Heritage Lottery Fund.

Putnam, Robert (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Rowe, N., Beasley, N., Adams, R., Long, J., Kay, T., Collins, M., Kirk, D., Coalter, F., Gratton, C., Taylor, P., Elson, M. & Ravenscroft, N. (2004). Driving up participation: The challenge for sport. London: Sport England.

Seaman, B. A. (2005). Attendance and public participation in the performing arts: A review of the empirical literature. Georgia State University, Andrew Young School of Policy Studies, Working Paper 05-03.

Sterman, J.D. (1987). Introductory comments to Archive section. In Forrester, N., The role of econometric techniques in dynamic modelling: systematic bias in the estimation of stock adjustment models. *System Dynamics Review* (3(1), 45-46).

Stigler, George J., & Becker, Gary S. (1977). De Gustibus Non Est Disputandum. *American Economic Review* (Vol. 67, No. 2, pp. 76-90).

Sullivan, O., & Katz-Gerro, T. (2007). The omnivore thesis revisited: Voracious cultural consumers. *European Sociological Review* (23:2, 123–137).

Throsby, D. (2001). *Economics and culture*. Cambridge: CUP.

Throsby, C. D., & Withers, G. (1979). *The economics of the performing arts*. New York: St Martin's Press.

Tisdell, C. (2006). *The economics of leisure: Volume I.* Cheltenham: Edward Elgar.

Veblen, Thorstein (1899). *The theory of the leisure class.* New York: New American Library.