Predictive factors from age 3 and infancy for poor child outcomes at age 5 relating to children's development, behaviour and health: evidence from the Millennium Cohort Study

John Hobcraft and Kathleen Kiernan
Department of Social Policy and Social Work
University of York

Contents

- **Overall Predictive Power of the Age 3 Factors**

7. What extra explanatory power can be derived from MCS1 measures?
- Behaviour scores
- Foundation Stage Profiles
- Child Health

8. Predictors of any poor outcome

9. Summary and Conclusion
- Changing Family Experiences from Infancy to Age 3
- Commonalities in retained predictors across outcomes at age 5
- Differences in retained predictors by outcome at age 5
- Outcomes at age 3
- Persistent Predictors
- Predictive performance
- Addition of birth/ age 9 month information
- Additional leverage from birth/ age 9 month information
- Experience of any of the three poor outcomes at age 5

10. References

1. Introduction

This project builds on an earlier project using data collected in the Millennium Cohort Study that identified factors during pregnancy and shortly after birth that were associated with child outcomes at age 5 years (Kiernan and Mensah, 2009). This study had a broader remit to identify family and individual level factors that help or hinder children's future well-being and health at age 5 drawing on the information collected on the Millennium Cohort children and their families at the 3 years old surveys supplemented by the 9 month survey.

It has the following foci:

1. Firstly, the identification of the attributes of the child and their family at the time the child was age 3 years that signal positive/negative outcomes for the child at age 5 years. A particular and new focus is on parenting behaviours and attitudes.

2. Secondly, it combines the information from age 9 months and age 3 years to assess the saliency of characteristics and their timing in childhood for the child outcomes.

3. Thirdly, it identifies families whose situations have improved (for example, moved out of poverty) or deteriorated (mother became depressed) between when the child was aged 9 months and age 3 years in order to assess how improvements or a deterioration in family well-being relates to the child outcomes.

The outcomes examined included how the children were doing on the Foundation Stage Profile carried out by their teachers in their first year at school; the extent to which they were exhibiting behaviour problems at age 5; and a measure of their health status. This was an ambitious and extensive exercise which examined a very large number of parental, family and child characteristics from pre-pregnancy, pregnancy, infancy and early childhood that might throw light on which children were faring less well at age 5 years. Our analytic strategy, which included extensive multivariate analyses, identified the salient variables relating to the child outcomes with the aim of informing the PREview tool.

2. The Millennium Cohort Study

The Millennium Cohort Study (MCS) is a nationally representative, large-scale longitudinal survey of children in the United Kingdom (Dex & Joshi 2005). For this analysis only children born in England are included. The first sweep of English families contained information from 11,533 families, including 11,695 children aged between 9 and 11 months but most of the children were aged 9 months. These children were born between September 2000 and August 2001 and thus would officially begin attending primary school in 2005.

The families have been followed up when the child was aged 3 and 5 years with achieved response rates of 78 and 79 per cent of the target sample respectively. Detailed information on the sampling strategy and response rates for the surveys can be found in Hansen (2008). Additionally, details on the survey, its origins, objectives, sampling and content of the survey waves are contained in the documentation attached to the data deposited with the UK Data Archive at Essex University.

The MCS sample design allowed for over-representation of families living in areas with high rates of child poverty or high proportions of ethnic minorities, which increased the power of the study to describe effects for these groups of families. The analyses presented take account of the initial sampling design as well as adjustments for non-response in the recruitment of the original sample and sample attrition over the follow up period to age 5. The study results are thus broadly representative for England (Plewis, 2007, Ketente, 2008).

The sample for this study was 8430 children born in England whose families participated at the age 5 survey. For analytic reasons the sample was restricted to one child per family.

2.1 The outcome measures

Three outcomes were considered relating to different dimensions of child development and well-being. The children's learning and development was assessed via the Foundation Stage Profile assessments carried out by teachers, which have been linked into the MCS records; their behaviour was assessed from responses to the Strength and Difficulties Questionnaire by mothers (or main carers) and their health status was also based on mother's (or main carer's) reports.

Children's developmental progress: Foundation Stage Profile

The Foundation Stage Profile (FSP) is an assessment of children's developmental achievement over the first year of primary school, assessing the Early Learning Goals for the children between ages 4 and 5 (Qualifications and Curriculum Authority 2003). The assessment was completed by the child's teacher for six areas of learning: personal, social and emotional development; communication, language and literacy; mathematical development; knowledge and understanding of the world; creative development; and physical development. The assessment includes continued observation over the year period and the assessments are routinely moderated. These data are collected for children in English state schools by the Department for Children, Schools and Families (formerly known as the Department for Education and Skills). The MCS survey data were linked to FSP assessments made over the academic year from 2005 to 2006, with a success rate of 95% for the cohort children attending state schools in England (Hansen & Jones 2008). The FSP aims to provide a rounded picture of a child's progress and development within their usual educational setting, and is appropriate for children of all abilities and children for whom English is an additional language (Qualifications and Curriculum Authority 2003). This assessment may be preferable to survey assessments which are usually made under test conditions, and which may be less appropriate for children of very low ability or for children with limited fluency in English. In this study we compare the experiences of children in approximately the bottom ten per cent of the distribution with the rest.

Children's Behaviour: Strength and Difficulties Questionnaire

At the age 5 interview the children's mothers or the main carer completed the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997). This is a 25 item behavioural screening questionnaire covering 5 different dimensions of children's behaviour: conduct problems, (often fights, often has temper/tantrums, generally obedient, argumentative with adults, can be spiteful to others); inattention-hyperactivity (restless/overactive, constantly fidgeting, easily distracted, can stop and think out before acting, sees tasks through to end), emotional symptoms (often unhappy, often complains of headaches, many worries, nervous or clingy, many fears), peer problems (rather solitary, tends to play alone; has at least one good friend; generally liked by other children; picked or bullied by other children; gets on better with adults than other children), and pro-social behaviour (considerate of other people's feelings; shares readily with other children; helpful if someone is hurt, upset or feeling ill; kind to younger children; often volunteers to help others). Each attribute was rated by the mother using a scale from 0 to 2 (not true, somewhat true, and certainly true) and coding was reversed for the positive attributes within the difficulties scales. Responses were summed to provide a total score for each

dimension. The first 4 scales were combined to yield a total behaviour problem score. A score of 17 or more across the population of children of all ages is regarded as a high score and typically identifies 10 per cent of children, although it varies according to the ages of the children (Goodman, 1997 and 2001). In this study the closest cut-off point to include children in the highest decile of difficulties was a score of 14 or more which captured 9 per cent of the children. We compared this group with the rest of the children with lower scores.

Children's Health Status

For this study we use a measure of overall health status where the child's health status at age 5 years was categorised by the mother as being excellent, very good, good, fair, or poor. This measure has been shown to correlate strongly with more direct measurements of health including children's chronic conditions and episodes of hospitalisation (Case et al, 2002). In the MCS sample 4 per cent of the mothers reported that their child was in fair or poor health. It is also worth noting that this self reported measure of health has been shown to be a reasonably good indicator of social inequalities in health when compared with more direct health measures (Subramanian and Ertel, 2008).

2.2 Age 3 Characteristics

The characteristics and experiences of the children's families at age 3 were divided into six main groupings:
- Parenting attitudes and behaviours;
- Partnership status and relations
- Mother's mental and general health;
- Socio-economic situation at the 3 year old survey;
- Demographic characteristics; and
- Age 3 child outcomes

The derivations of these variables are described below and the sample distributions are shown in Tables 1-6. We describe all the factors that were considered for inclusion in our analyses although not all remained in our final models. As we proceed we highlight the factors that were **not** retained in our final models. These were very few in number and included drinking behaviours, use of recreational drugs, an indicator of domestic violence and the SEG status of the mother's and the father's/partner's occupation.

2.3.1 Parenting attitudes and behaviours

A range of information was collected at the 3 year old survey on different aspects of parenting. We have categorised these into six groups: the mother's relations with the child in terms of whether the relationship in her view is warm or conflictual and whether parent-child interactions observed in the home by the interviewer were positive or negative; those relating to the promotion of the child's cognitive development; family organisation; disciplinary practices; and partnership disagreements in relation to the child.

Maternal Relations with the child: Pianta Scale

The main respondent was asked a series of questions about their relationship with the child at the 3 year survey using the Pianta child parent relationship scale (Pianta 1995). Seven items were selected that described the level of warmth in the relationship and 7 that described the level of conflict in the relationship.

For warmth the items were:

> *I share an affectionate, warm relationship with the Child*
> *Child will seek comfort from me*
> *Child values his/her relationship with me*
> *When praised Child beams with pride*
> *Child spontaneously shares information about himself/herself*
> *It is easy to be in tune with what the Child is feeling*
> *Child shares his/her feelings and experiences with me*

For Conflict the items were*:*

> *Child and I always seem to be struggling with each other*
> *Child easily becomes angry with me*
> *Child remains angry or is resistant after being disciplined*
> *Dealing with the Child drains my energy*
> *When the Child wakes up in a bad mood, I know we're in for a long and difficult day*
> *The Child's feelings towards me can be unpredictable or can change suddenly*
> *The Child is sneaky or manipulative with me*

Each item was scored as 1 definitely does not apply, 2 not really, 3 neutral, 4 applies sometimes, or 5 definitely applies. Can't say responses were considered as missing information. Scores were summed for parents who had completed all warmth items and grouped as 35 (highest warmth), 33-34, 30-32, or 7-29 (low warmth). Scores were summed for parents who had completed all conflict items and grouped as 7-15 (lowest conflict), 16-20, 21-26, or 27-35 (high conflict).

Relations with child: Interviewer Observations

As well as information from the mother on the warmth or otherwise of her relations with the child we have some more independent information based on observations by the interviewer. The interviewer recorded whether they had observed the following positive interactions between the parent and the child during the cognitive assessment task:

>*Parent's voice positive when speaking to the child*
>*Parent converses at least twice with child*
>*Parent answers child's questions verbally*
>*Parent praises child spontaneously*
>*Parent caresses or kisses child*
>*Parent introduces interviewer to the child*

These were grouped according to the number of positive behaviours observed as 6, 5, 3-4, or 0-2.

The interviewer was also asked to record whether they had observed the following negative behaviours by the parent towards the child during the cognitive assessment task:

>*Parent scolded the child more than once*
>*Parent used physical restraint on the child*
>*Parent slapped or spanked child*

These were grouped according to the number of negative behaviours observed as 0, 1, or 2-3.

Promotion of reading and learning

The main respondent was asked how frequently they read to their child at the 3 year survey, selecting from the categories, every day, several times per week, once or twice per week, once or twice per month, or less often/ not at all. A high proportion of parents read to their child either daily or several times a week (79 per cent).

Family organisation

At the 3 year survey the main respondent was asked whether the child goes to bed at regular times, selecting from always, usually, sometimes or never/ almost never and was also asked whether the child has meals at regular times, selecting from always, usually, sometimes or never/ almost never. These variables were dichotomised into frequent (the always and usually categories which) and not frequent (the sometimes and never or almost never categories). Overall 9 per cent of the children had irregular mealtimes and 20 per cent had irregular bedtimes.

Disciplinary practices

Information on was collected on whether the child was smacked or shouted at when they were naughty. For this analysis whether parents shouted at the child was grouped into two groups frequent and not frequent which compared daily shouting with the rest. Seventeen per cent of the parents reported that they shouted at their child every day. Smacking which was reported on a much less frequently was grouped as not frequent and frequent in which frequent included those who responded that they smacked their child once a week or more or daily (9.5 per cent of the parents). Smacking did not enter our final models.

Partnership Disagreement over the child

Information was also collected on how often the parents disagreed over issues concerning the cohort child. The range of answers ranged from never, through less than once a week, once a week, several times a week, daily to more than once a day. Only 6 per cent had disagreements several times a week or more so the last three categories were combined.

2.3.2 Partnership Status and Relations

The family status measure included four groups of families based on their situation at the time of the 3 year old survey. These were: biological parents who were married to one another (69 per cent); biological parents who were cohabiting (15 per cent); lone mother families (14 per cent) and step families (2 per cent). In this latter group the father was not the biological father and in two out of three of these cases the mother and social father were cohabiting.

Domestic violence and parental conflict are important factors that may affect children's well-being. There is very limited information on this in the 3 year old survey. There was a question which was posed in the following way: "People often use force in a relationship – grabbing, pushing, shaking hitting kicking etc... Has your husband/partner ever used force for any reason yes or no? The percentage answering yes was very low at 4 per cent. This variable did not enter our models which may be due to the small numbers involved but disagreements over the child did, which may represent an element of parental conflict within the family.

2.3.3 Mother's Mental and General health

Mental health

Two indicators of maternal mental health were included in our analyses. The first is derived from information on whether a doctor has ever diagnosed depression or serious anxiety and whether the mother is currently receiving treatment for such conditions. Seventy one per cent of the mothers reported that they had never been diagnosed with depression or severe anxiety, 21 per cent had been diagnosed in the past but were not being currently treated and 8 per cent were currently being treated.

Our second indicator was measured using the Kessler 6 scale [Kessler et al 2002] which was administered as a computerised self report to parents who were resident in the household at the age 3 survey. The Kessler 6 scale provides a measure of psychological distress from the respondent's report of how often over the last 30 days they had felt depressed, hopeless, restless or fidgety, that everything you did was an effort, worthless, and nervous. For each item the respondent indicated whether they have felt this way none, a little, some, most, or all of the time which are scored from 0 to 4 respectively. The questions form a 24 point scale and for this study scores were grouped as 0 to 3, 4 to 6, 7 to 12 and 13 to 24 indicating a gradient in levels of psychological distress.

The Kessler scale has been evaluated as a screen for prevalence of serious mental illness within a community population of US adults. In this evaluation a score of 13 or more was indicated as an appropriate reference level to estimate the prevalence of serious mental illness in the population (Kessler et al 2003). Amongst responding mothers 3.2 per cent reported symptoms at this level. At a less stringent definition, 7 to 24 points, 14.2 per cent of the mothers could be identified as experiencing high symptom levels.

General Health

As well as information on mental health mothers also reported whether their general health was excellent, good, fair or poor. The responses for the initial analysis were sub-divided into three groups excellent, good and a report of fair or poor health was used to indicate health difficulties. Eighteen per cent of the mothers reported having fair or poor health.

Drinking, Drugs and Smoking around the Child

Information on drinking habits and use of recreational drugs was collected from the main respondent. For our analysis frequency of drinking was divided into 5 categories: never (19 per cent of the sample), rarely (36 per cent), once or twice a

week (26 per cent) three to four times a week 12 per cent) and five times or more a week (8 per cent). Only a small minority (3.5 per cent) reported that they used recreational drugs. With reference to smoking around the child information was collected on whether anyone smoked in the same room as the child and 16 per cent of the replies were in the affirmative. Of these factors only smoking in the same room as the child was retained in our final models.

2.3.4 Socio-economic situation at the 3 year old survey

In the MCS study socio-economic situation was measured in a variety of ways and in this study we included information on household employment and socio-economic status, financial status, housing tenure. We also included information on the index of multiple deprivation for the area in which the child lived. The level of educational qualifications attained is also described in this section.

Worklessness and Socio-economic group
We used a simple indicator of household employment namely whether anyone in the household was in employment or not.   Fifteen per cent the children lived in workless households.
For the families where parents were in work the occupations of the mother and the father/partner were divided into three Socio-economic groupings: managerial/ professional; intermediate and small employers and non-professional self employed; and routine and technical. Socio-economic group was not retained in any of our final models. .

Income and poverty measures
The family income of the household was reported by selecting a net household income band from a show card. We grouped the incomes into five bands broadly representing the quintiles of the distribution. The incomes were equivalised using the OECD criteria (ref). We also included two poverty measures in our analysis: an income poverty measure and a benefits poverty measure. The family was deemed to be living in poverty if the equivalised household income was below 60 per cent of the median before housing costs and on this basis 26 per cent per cent of children were living in poor families at the time of the 3 year old survey.

The benefits poverty measure, which distinguishes families in receipt of means-tested benefits, is a useful indication of families living in poverty in the absence of income data. Eighteen per cent of the families were deemed to be poor on the benefit measure i.e. they were receiving Income Support or were in receipt of Working Family Tax Credit or Job Seeker's Allowance as well as receiving Housing Benefit or Council Tax Benefit.

Housing tenure, housing difficulties and Index of Multiple Deprivation (IMD)
The tenure of the family home was classified as owner occupier (67per cent), privately rented (7 per cent), social housing including renting from a local authority or housing association (22 per cent), or other type of tenure which for the majority of families represented here were living with their parents.

We also included a measure of housing difficulties derived from a number of items relating to homelessness, and whether they had had to move because they could no longer afford their home, had experienced eviction or repossession, or had problems with neighbours: on this basis 4 per cent of the families had housing difficulties.

The 2004 Index of Multiple Deprivation (Office of the Deputy Prime Minister, 2004) which combines a number of indicators, chosen to cover a range of economic, social and housing issues, into a single deprivation score for each small area in England has also been linked into the MCS Survey data. For this analysis the scores for the area in which the child resided at age 3 were used and have been broadly divided into quintiles.

Mother's qualifications
Educational qualifications represent another aspect of socio-economic status, although probably having wider implications too. Parents' academic or vocational qualifications were reported at the first survey and updated at the follow up surveys. These were classified into 5 levels equivalent to the National Vocational Qualification scale (NVQ) ranging from no qualifications to level 4 or 5 representing professional or academic qualifications at degree level, nursing or other medical qualifications. For our analyses mother's qualifications attained by the time of the age 3 survey were reduced to four groupings with NVQ 4 and 5 being combined into one group.

2.3.5 Demographic Characteristics

A number of what can be broadly termed demographic factors were included in our models. These included the child's gender, whether they were the first born child and mother's age at first birth.

Age of the child in months was tested for inclusion in all models as the children were born at different times of the year which may be important in relation to how well they are doing on, for example, the Foundation Stage Profile. For the analysis the months of birth were grouped into 4 groups: September to December; January to March; April to June; and July and August.

We included a new variable that was collected retrospectively at the 3 year old survey on whether the mother of the child had been born in the UK or not. Twelve per cent of the mother's reported that they had been born outside the UK.

Information on ethnicity of the child was also considered in our analyses. The main respondent was asked which ethnic identity category they felt their child belonged to, corresponding to the categories used in the UK census [ONS, 2003 39]. Preliminary examination of the data showed that the most appropriate groupings for the analysis, given constraints of sample size, was to divide the child's ethnicity into three groups: White (85 per cent); Pakistani or Bangladeshi (5 per cent) and the rest of the ethnic groups (10 per cent).

As well as these more fixed attributes three others relating to the situation at age 3 were included these were: the number of children in the family; whether a grandparent lived in the household (4 per cent of the families); and language usually spoken in the home, which was grouped into English only, English plus another language and another language only.

Another variable that does not fit readily under our headings which was collected retrospectively at the 5 year old survey was whether the child had used any early care or education facility prior to starting school. We can't give the exact timing of the attendance but our exploratory analyses showed that non attendance was an important variable in relation to some poorer child outcomes. This was a small minority of children: 7 per cent.

2.3.6 Age 3 Child outcomes

The focus of this study is on the child's cognitive, behavioural and health situation at age 5. However, how a child is doing at age 3 across these domains is likely to be strongly related to how they are doing at age 5. For the MCS children we have measures on how they were doing across these domains at age 3. The measures differ from those at age 5 with the exception of the SDQ which was collected at both ages. Our results on what age 3 factors matter for the child outcomes at age 5 include models with and without the age 3 outcomes.

Cognitive

At the three year old interview the children's stage of cognitive development was assessed via six tests of the *Bracken Basic Concept Scale* (BBCS) administered to the children, which assessed comprehension of: colours, letters, numbers, sizes, comparisons of objects and shapes; these tests provide an indication of the child's readiness for formal schooling (Bracken, 2002). The raw scores from these tests were

added up, normalised and grouped into 5 categories: very delayed, delayed, average, advanced, and very advanced.

Behaviour

Behavioural adjustment had also been assessed at age 3 with the Strengths and Difficulties Questionnaire. For the analyses where the age 3 child outcomes are included in the models the scores were divided into three groups differentiating those in the top 10 per cent of scores (16-30) with those with an intermediate score (12-15: 17 per cent) and low scores 0-11.

Long-standing disability or illness

Information on the child's health was derived from the following questions. "Does the child have long term (health) conditions that have been diagnosed by a health professional? By long term I mean anything that the Cohort child has had for at least 3 months or is expected to continue for at least the next 3 months. If so - does this limit him/her at play or from joining in any other activity normal for a child his/her age?" In the sample 13 per cent of the children had a longstanding illness that did not inhibit play and activities, and 3 per cent one that did.

2.4 Explanatory Factors from the 9 month old survey -update

In the first report, Kiernan and Mensah (2009) examined the maternal indicators during pregnancy and the child's infancy that were related to the child's development, behaviour and health at age 5. The earlier report and the analyses therein focused on factors that had the potential to be collected during pregnancy. Our remit in this report is wider and we include additional measures, some of which were available at the time of writing the first report such as mother's attitude to her child's development, whether the mother had a limiting longstanding disability, and whether the child was living in a workless household. Other data have become available since the first report was written. This includes information on the mother's Body Mass Index (BMI) pre-pregnancy that allowed us to estimate whether the mother was overweight or obese pre-pregnancy. The mother's attitude to child development at age 9 months was derived from her level of agreement with four items that were concerned with how a baby should be treated. These were the importance of developing a regular pattern of feeding and sleeping; the extent to which babies need to be stimulated; and the extent to which she viewed talking and cuddling a baby as being important.   The answers to these questions were summed to produce a scale and coded into groups where one represented the most positive attitudes.   These variables and their distributions are shown in Table 7. A full list of all the variables used from the 9 month old survey is listed in Appendix A.

3. Methods

As outlined above, there are a very wide range of measures collected at age 3 in the MCS; these can be supplemented by a similarly wide range of measures from birth/ age 9 months. Any analysis that included all of these multiple indicators would be extraordinarily complex and make it impossible to distinguish the wood from the trees. Our goal is thus to explore ways of simplifying the amount of information used whilst as far as possible retaining maximal explanatory power. We begin this process with a descriptive analysis of the associations of each of the three outcomes at age 5 with each measure at age 3 to provide an informative backcloth. We then move on to more elaborate regression analyses that aim, in a systematic way, to retain only those key predictors that are important or sometimes to retain one which has slightly more powerful among two or more close competitors. Since the measures of similar outcomes at age 3 to those being predicted for age 5 are expected to be enormously powerful predictors of the outcomes at age 5 we consider models that exclude and include these measures. We then consider the predictive power of the selected models. Finally we further explore which measures from the survey at age 9 months retain predictive power even when the most powerful associations at age 3 are known.

3.1 Bivariate description

Initially we provide a bivariate description of the relationship of each measure at age 3 to each of the outcomes at age 5. These associations are presented in groups by domain for the age 3 measures: parenting, partnership, maternal health, socioeconomic situation, demographic characteristics, and outcome measures at age 3. For each separate measure the distribution across the sample that have valid responses on that measure is shown. The bivariate association is assessed using a logistic regression model, adjusted for the sample design and attrition, to provide the 'raw' odds ratios for each of the three outcomes at age 5. In addition the p-values associated with contrasts of these odds ratios with the reference group (usually the most advantaged group) are shown.

Thus, for example, in Table 1 we see the first measure, the parental warmth as measured by the Pianta scale. Nearly half the sample have the maximum score of 35 representing that each of the 7 statements about parental warmth 'definitely applies'. A further 30 per cent have scores of 33-34, still very high; and another 15 per cent have moderate scores of 30-32; lastly about five per cent have 'low' scores, below 30, covering a broad range.

Despite the range of parental warmth responses being fairly narrow, we see quite substantial differentials in child outcomes at age 5 by this measure. For the learning and development FSP scores at age 5, we see only a small and statistically insignificant association with the slightly reduced warmth measure, but the 'moderate' warmth has an odds ratio of 1.83:1 for poor learning and development compared with the 'maximum' reference group and the 'low' group has an odds ratio of 3.77:1; both are very strongly statistically significant.

For poor behaviour scores at age 5, measured by the SDQ, we see very strong associations with parental warmth, with even the slightly reduced warmth has an odds ratio of 1.73:1, the moderate warmth category has odds of 2.36:1 and the low group of 6.44:1. There are also quite strong associations with child health at age 5, with odds ratios being 1.46, 1.88, 1nd 2.47 respectively for the three groups; levels of significance are slightly less powerful here because the prevalence of the poor health outcome is lower.

A further section provides a descriptive analysis with bivariate odds ratios of combinations of family status on the same measures of disadvantage at age 9 months and age 3. These measures are used to classify the sample into those who are 'persistently' deprived (at both waves), those who were 'episodically deprived', either improving or worsening from age 9 months to age 3, and those who were not deprived by the measure at either wave. The domains explored encompass poverty, employment status, maternal depression and maternal general health. In addition a similar analysis looking at those children whose parents split up between age 9 months and age 3 is included.

3.2 Multivariate modelling strategy

In the subsequent sections we present several multivariate logistic regressions for each of the three outcomes at age 5 (and for a combined 'any poor outcome' measure). Our first chosen model represents the set of measures from the age 3 survey, excluding the parallel outcome measures at age 3. We then explore the additional predictive power of including these age 3 outcome measures and choose a 'best' model incorporating these additional elements. This is followed by finding out which additional measures taken from the age 9 month survey added further predictive power to the model that included all of the retained measures from the age 3 survey.

Several steps were implemented in the choice of which predictive factors should be retained in the model that considered the age three measures, excluding the outcome measures at age 3. All predictors are coded categorically and missing values

are coded as a further category. The initial exploratory steps involved fitting a stepwise logistic model using forwards selection and another using backwards selection; this procedure was repeated both for the individual dummy variables for each category and for the introduction (or exclusion) of each variable as a whole. The four resultant models were then compared and usually found to overlap very considerably. The measures that remained in any of these were considered further and those that were not included in any of these models were temporarily dropped from the analysis. A preliminary final model was then found through careful evaluation of the marginal indicators, screening through checking statistical significance using Wald chi-square tests in models that adjust for the survey design and attrition. Where the combination of two (or more) categories on a single variable were jointly significant and had similar odds ratios, these categories were pooled to simplify the model. Once this cleaning up had been done a further sweep of checking whether any of the temporarily excluded variables (or temporarily omitted categories of included variables) using Wald tests was done and a final model chosen that only included those measures that added significant explanatory power.

 A similar process was followed to choose a 'final' model that drew solely on information collected at the first wave of the MCS, when the cohort members were aged 9 months. The addition of the age 3 outcome measures to the age 3 'final' model involved exploring both whether the three age 3 outcomes and their respective categories were indeed powerful enough predictors to be retained and an assessment of which factors that had been retained from the age 3 analysis might be removed from the enhanced model without loss of power. Again this was checked thoroughly. Finally, we asked which factors from the 'final' age 9 month model added explanatory power to this enhanced model that drew upon all of the information at age 3 including the outcome measures. Once again, any age 3 predictors that became redundant after inclusion of key predictors from birth/ age 9 months were removed and carefully checked. Where statistically justified the models retained parameters corresponding to missing values for a measure (in order to retain maximal sample size), but the odds ratios for these are not presented in the tables, since they are not of substantive interest.

All of the models presented were computed using the Stata software package and using logistic regression with the set of 'survey' commands that adjust for the complex sampled design of the MCS and includes adjustment for attrition too. Results are presented as Odds Ratios and p-values are provided for the contrast with the relevant (omitted) reference group.

3.3 Judging the predictive power of the models

Summary statistics on overall goodness of fit are notoriously hard to interpret, especially where the regressions allow for complex survey designs as here. For such models the main information is provided by the F-statistic, shown in our Tables, in relation to the degrees of freedom in the model. Since our models predominantly retain factors that have individually statistically significant associations with the outcomes it is no surprise that the chosen models are significantly better than the null model. We also present a Wald chi-square statistic, although again comparisons between models are not valid given that successive models are not nested. A fairly widely used overall measure for model selection is the Bayesian Information Criterion (BIC) which is an adjusted chi-square statistic that is penalised for model complexity (see Raftery 1995); generally models with a lower BIC are to be preferred to those with a higher one and the penalisation for additional degrees of freedom avoids unnecessary clutter. However, this statistic is not available for complex survey models, so we have provided an approximation from running the same logistic models without the corrections for survey design.

Perhaps the most useful summary statistic for the predictive power of the models is the Receiver Operating Characteristic (ROC) which is widely used in medicine, radiology and other fields and closely related to the Mann-Whitney U test, the Wilcoxon test of ranks and the Gini coefficient used by economists to measure income inequality. Where an outcome is dichotomous, as here, there is always a trade off between sensitivity, the proportion of actual cases that are identified – true positives, and specificity, the proportion of non-cases that are correctly identified – true negatives. The ROC curve is a graph of the sensitivity against 1-specificity for a binary classification; alternatively it compares the true positive rate (sensitivity) against the false positive rate (1-specificity). These measures vary as the discrimination threshold is varied and capture in continuous form the distributions shown for predictive power in the previous report (Kiernan and Mensah 2009). Moreover, it is possible to test whether different models have significantly better predictive power using the summary statistic of the area under the curve.

4. Bivariate Description

Tables 1 to 6 show how the age 3 characteristics relate to the child outcomes both in terms of their distributions also the results from logistic regression analyses expressed in terms of odds ratios which allow us to assess the importance of the individual factors. Table 7 provides the analogous information for the variables from the 9 month old survey that were not discussed in the earlier report.   These are provided for background information and we will only present some highlights from the Tables as they are fairly self-explanatory.

4.1 Age 3 factors and Childhood outcomes

It is clear from Table 1 that virtually all the parenting measures are strongly associated with the child outcomes. The exceptions are that negative home observations and shouting at the child are not significantly associated with poor child health and frequent smacking is strongly associated with behaviour problems but has little association with the cognitive and health outcomes. Some of the associations are quite strong. For example, the odds of child having high levels of behaviour problems at age 5 are 16 times greater if the mother had a conflictual relationship with the child at age 3 compared to where the mother had low conflict relationship.

Table 2 shows the associations between family structures at age 3 and use of force in the family at this age and how they are associated with our child outcomes. Use of force is negatively associated with the child outcomes but perhaps due to the low reported incidence this did not enter into our final models.

From Table 3 we see that maternal depression and mother's general health are important factors in distinguishing between children who are not faring well cognitively, behaviourally and health wise.   Exposure to adult smoking is also associated with all the outcomes but the associations with the frequency of drinking alcohol are less clear cut. Children in the group where the mothers reported that they never drank tended to have the highest odds of doing less well on all the outcomes. This may be due to the fact that women who drink are more likely to be drawn from the higher socio-economic groups (Dex and Joshi, 2005) and as we see in Table 4 their children on average have fewer difficulties than women from lower socio-economic groups. Use of recreational drugs is associated with child behaviour problems but not how they were doing at school or in the health domain. Both drinking and use of recreational drugs did not enter into our final models.

Turning our attention to the socio-economic characteristics of the children's families it is clear from the data shown in Table 4 that nearly all the factors are strongly related to the child outcomes. Parental unemployment, social status of the occupation of the mother or the father/partner, level of household income, living in poverty, being in receipt of means-tested benefits, living in rented accommodation, experience of housing difficulties, and living in a relatively deprived area are all associated with negative child outcomes. Level of education attained by the mother is one of the strongest influences. Many of these factors are inter-related and this will be taken into account in our final models.

It is clear from Table 5 that boys fare less well than girls on all the outcomes. The timing of the child's birth is particularly important in relation to the Foundation Stage Profile assessment and whether the child's mother was born in the UK is also important in relation to this outcome, whereas a child's ethnicity and whether languages other than English are spoken in the home tend to be associated with all the outcomes. Whether a grandparent co-resides with the child, having a greater numbers of siblings and not having attended pre-school care or education are all negatively associated with the outcomes.

As might be expected how children were doing cognitively, behaviourally and health wise at age 3 was strongly related to how they were doing in the same domains at age 5. This can be clearly seen in Table 6 as well indications that children doing less well in one domain at age 3 were also likely to be doing less well in other domains at age 5.

4.2 Additional factors from age 9 months survey

The distributions and odds ratios for the additional 9 month old variables are shown in Table 7. With regard to the mother's BMI prior to pregnancy we see that where mothers were either under weight or morbidly obese their children were less likely to be doing well at school compared with mothers in the normal BMI range and children of underweight mothers were also more likely to have behaviour problems. Slightly overweight mothers were also more likely to report behaviour problems and health problems amongst their children. Whether the mother had a limiting longstanding disability and whether there was no one working in the household at the 9 month old survey was also associated with the three outcomes. We also see that the less positive the mother's attitude was to her child's development the more the child was likely to be doing less well at school and the more likely they were to have behaviour problems at age 5. But mother's attitude to her child's development was only weakly related to the child health outcome.

5. Changing Family Experiences from Infancy to Age 3

The aim of this section is to examine whether changing circumstances in terms of improvement or deterioration in family situations is related to how the children are doing in terms of their performance on Foundation Stage Profile, whether they are exhibiting behavioural problems and whether they are in fair/poor health at the time of the age 5 survey. We examine changes in the poverty status and employment status of the families; whether the child's parents had separated; and the health of the mothers in terms of her general health and whether she was suffering from depressive symptoms.

5.1 Poverty and Employment status

A family was deemed to be living in poverty if the household income was 60 per cent below the UK median before housing costs. On this measure 25 per cent of the children were living in poverty at 9 months and 26 per cent were living in poverty at age 3. Seventeen per cent of the children lived in poverty on both occasions (which we will refer to as persistent poverty but recognising that some of these families may have moved in and out of poverty in the interim period), 7 per cent were in poverty only at the 9 month old survey and 9 per cent at the time of the three year old survey (these two groups will be referred to as experiencing episodic periods of poverty).

The employment status of the family also changed, with some households having parents who moved in or out of employment. Given the extent of lone parent families (13 at the 9 month old survey and 18 per cent at the 3 year old survey) we used a classification of whether there was one parent in work or not. This parent was not confined to biological parents so for example if there was a social father in the household his work status was also included. Across the two surveys 80 per cent of the families had someone in work and 10 per cent were workless households. The remaining 10 per cent of families had no one in employment at one or other of the surveys made up of 5 per cent of families at the 9 month old survey and a similar proportion at the 3 year old survey.

The poverty and employment measures are of course likely to overlap but they are by no means co-incident. So for example, one in five (19 per cent) of the families who had not experienced poverty at the time of the 9 month old and 3 year old surveys had experienced unemployment on one or other occasion. Moreover, in terms of a Preview tool it may be easier to collect information on employment than income and looking to the future it may be possible to link employment records from administrative data.

Table 8.1 shows the proportions of children living in families with differing economic situations and the extent to which children in these different family settings were to be found in the bottom deciles of the behaviour problems inventory or the Foundation Stage Profile , or were reported to be in poor or fair health. We also provide information on the percentages falling into any of these categories. Table 8.1 also show the odds of children in the various settings being different from the reference group of children with no recorded experience of poverty or unemployment in their families. This enables us to assess the extent to which episodic or persistent experience matters; and also to assess the extent to which improvements or deterioration between 9 months and age 3 matters.

Table 8.1: Relationship between changing poverty and work status from 9 months to age 3 and child outcomes at age 5

	FSP lowest decile			SDQ highest decile			Health fair/poor			Any poor outcome	
	9.5%			8.5%			3.8%			16.8%	
Poverty status	%	Odds ratios		%	Odds ratios		%	Odds ratios		%	Odds ratios
None	4.9	1.0 (ref)		5.2	1.0 (ref)		2.3	1.0 (ref)		10.5	1.0 (ref)
9 months only	13.3	2.9 p=0.000		13.1	2.8 p=000		4.6	2.1 p=0.002		23.5	2.6 p=0.000
Age 3 only	12.2	2.7 p=0.000		11.4	2.4 p=0.000		5.4	2.5 p=0.001		23.3	2.6 P=0.000
Both surveys	20.0	4.8 p=0.000		17.6	3.9 p=0.000		7.1	3.3 p=0.000		32.1	4.0 p=0.000
Work status											
No unemployment	6.6	1.0		6.3	1.0		2.9	1.0		12.9	1.0
Unemployment 9 months only	15.8	2.7 p=0.000		12.2	2.0 p=0.000		6.9	2.5 p=0.000		27.5	2.6 p=0.000
Unemployment age 3 only	20.6	3.7 p=0.000		18.0	3.2 p=0.000		5.4	1.9 p=0.006		34.5	3.6 p=0.000
Unemployment at both surveys	23.4	4.4 p=0.000		21.1	4.0 p=0.000		8.2	3.1 p=0.000		37.8	4.1 p=0.000

It is clear that there is an association between experience of childhood poverty and children's cognitive development with both episodic and persistent poverty taking

their toll. Children who have experienced persistent poverty (17 per cent of the sample) were over 5 times as likely to be in the bottom decile of the FSP compared with children who had not experienced poverty. Children who experienced poverty on one occasion whether in infancy or early childhood held an intermediate position with the proportions being in the bottom 10 per cent of the foundation stage profile being similar (14 and 13 per cent respectively). It is striking that already at this early age, one in five of the children in persistently poor households are exhibiting what could be regarded as cognitive delay and this was also found to be the case in our earlier work on age 3 outcomes (Kiernan and Mensah, 2009a). We see a similar picture for behaviour problems in that children who experienced poverty on one occasion whether in infancy or early childhood held an intermediate position and the proportions exhibiting behaviour problems were similar regardless of whether the poverty was experienced during infancy or early childhood. Again children in persistently poor families were substantially more likely to have behavioural problems than those in non-poor families (18 per cent compared with 5 per cent). Looking at the extent to which children are reported as being in fair or poor health we see a similar pattern with the persistently poor being more likely to be in the fair/poor category (7 per cent) compared with 2 per cent of children with no recorded experience of poverty at 9 months and age 3 years and 5 per cent of those who had experienced episodic poverty..

The findings with respect to unemployed families are broadly similar to those for poverty but there are some differences worth noting. Compared with the findings for poverty history we see that the legacy of worklessness is less if it occurred earlier rather than later which contrasts with the findings for poverty where there seems to be a similar legacy for early and later episodic poverty particularly with regard to the cognitive and behavioural outcomes. This is less the case for the child health outcome where the earlier episode has a stronger association.

5.2 Maternal Depression and General Health Status

We now consider the association between changes in maternal depression and maternal health for child outcomes. The measures of maternal depression for the two surveys were not identical. At the 9 month survey a reduced form of the Rutter Malaise Inventory was administered which contained 9 items and was described in detail in the first report (Kiernan and Mensah, 2009). At the 3 year survey as described earlier in this report the Kessler 6 scale was administered in which the mothers indicated how often they had experienced the 6 items over the last 30 days (Kessler, 2002 and 2003). For each item the respondent indicated whether they had felt this way none, a little, some, most or all of the time which were scored from 0 to 4 respectively, forming a 24 point scale. For this study following our earlier work

scores of at least 4 points were used to classify high levels of psychological distress (Mensah & Kiernan, 2009).

Mothers were grouped into those who did not report psychological distress at either survey (78 per cent), those who reported psychological distress at 9 months only (7 per cent), those who reported psychological distress at 3 years only (9 per cent), and those who reported psychological distress at both of the surveys (6 per cent). Again we use the nomenclature of episodic and persistent when referring to these groupings.

Table 8.2: Relationship between maternal depression and health status from 9 months to age 3 and child outcomes at age 5

	FSP lowest decile			SDQ highest decile			Health fair/poor			Any poor outcome	
	9.5%			8.5%			3.8%			16.8%	
Maternal depression	%	Odds ratios		%	Odds ratios		%	Odds ratios		%	Odds ratios
None	6.5	1.0 (ref)		5.0	1.0 (ref)		2.4	1.0 (ref)		11.7	1.0 (ref)
9 months only	11.4	1.9 p=0.001		15.2	3.4 p=000		4.7	2.0 p=0.004		24.1	2.4 p=0.000
Age 3 only	11.8	1.9 p=0.000		16.5	3.7 p=0.000		5.8	2.5 p=0.001		26.2	2.7 P=0.000
Both surveys	17.3	3.0 p=0.000		25.7	6.6 p=0.000		8.8	3.9 p=0.000		35.0	4.1 p=0.000
Health status											
Excellent/good both surveys	6.6	1.0		6.3	1.0		2.9	1.0		12.9	1.0
Fair/poor 9 months only	15.8	2.7 p=0.000		12.2	2.0 p=0.000		6.9	2.5 p=0.000		27.5	2.6 p=0.000
Fair/poor age 3 only	20.6	3.7 p=0.000		18.0	3.2 p=0.000		5.4	1.9 p=0.006		34.5	3.6 p=0.000
Fair/poor at both surveys	23.4	4.4 p=0.000		21.1	4.0 p=0.000		8.2	3.1 p=0.000		37.8	4.1 p=0.000

At each of the 9 month and 3 years surveys mothers reported whether their general health was excellent, good, fair or poor. A report of fair or poor health was used to indicate health difficulties at either of these times. Mothers were grouped into those who did not report health difficulties at either survey, those who reported difficulties

at 9 months only, those who reported difficulties at 3 years only, and those who reported difficulties at both of the surveys which we refer to as persistent health difficulties.

From Table 8.2 which provides information on maternal depression and maternal health we see that across all the outcomes there is a similar pattern. For example, children with persistently depressed mothers or mothers who reported that their health was fair/poor on both occasions were much more likely to be doing less well in school and more likely to have behavioural problems and to be in poorer health than their peers whose had not lived with depressed mothers or less healthy mothers. Some of the differences are quite large. For example, 27 per cent of the children in families where the mother was persistently depressed had a high level of behavioural problems as compared with 5 per cent of those whose mothers did not report depression on either occasion. The former group of children were also over twice as likely to be in the lowest decile of the FSP and three times as likely to have fair/poor health. Children whose mothers had experienced episodic depression or fair/poor health held an intermediate position and for both these maternal characteristics similar proportions were not doing well at school, behaviourally or health wise.

5.3 Stability and Change in Family Structures

Between the time the child was aged 9 months and 3 years 8.6 per cent of their parents had separated. Less of the married than cohabiting parents had separated; 6.5 per cent compared with 17.4 per cent. For this analysis we have combined the two groups of parents and examine the extent to which parental separation is associated with poorer outcomes amongst their children.

From Table 8.3 we see that children whose parents had separated were doing less well on the FSP and on the Strength and Difficulties measure. Twice as many children from the separated group compared with the intact group were in the lowest decile of these two measures. The relationship between parental separation and the health outcome was weaker but was still significantly different at the 5 per cent level.

Table 8.3: Relationship between parental separation occurring between 9 months and age 3 and child outcomes at age 5

	FSP lowest decile 9.5%		SDQ highest decile 8.5%		Health fair/poor 3.8%		Any poor outcome 16.8%	
Parental separation	%	Odds Ratio	%	Odds Ratio	%	Odds ratio	%	Odds ratio
No	7.8	1.0 (ref)	6.8	1.0 (ref)	3.2	1.0 (ref)	14.3	1.0 (ref)
Yes	17.0	2.41 p=0.000	14.0	2.24 p=0.000	4.9	1.55 p=0.044	29.1	2.45 p=0.000

Conclusion

It is clear from these analyses that there are legacies of episodic poverty, worklessness, maternal depression and poor health on children's well-being at age 5 but persistence of these attributes tend to be more deleterious.

6. Age 3 predictors for child outcomes

In this section we present the results of the extensive exploration of multivariate models including predictors from the MCS interview at age 3 for the three outcomes at age 5. The methods were described in section 3.2.

6.1 Selected key indicators

In Figure 1 we highlight the factors that, net of all the other factors, continue to have significant associations with the child outcomes at age 5. Two groups are distinguished: those factors that are retained even after controls for outcomes at age 3 are included; and those that were retained in the model that did not include the outcomes at age 3 but dropped once these were included. It is worth looking at the broad picture provided by this summary before examining the detailed results presented subsequently in Table 5 and Figures 6-8.

Whilst there is some similarity in the factors retained for the three different outcomes at age 5, what is perhaps more striking are the differences, suggesting that combining these three outcomes into a summary indicator of any poor outcome at age 5 is unwise. Beginning with the contrasts between the models for the lowest 9% on SDQ behaviour scores and the lowest 9% on FSP developmental scores, which retain similar numbers of factors, we see remarkable differences in the key predictors. These two models share gender of the child, ethnicity, mother's qualification levels and net household income as key predictors; although both ethnicity and net household income drop out of the model for FSP once the outcome measures at age 3 are included. On the other hand, the key indicators selected for the SDQ model include partnership status, a wide range of negative indications of parenting (Pianta conflict, irregular mealtimes, shouting at the child, disagreements with partner over child, and smoking around the child), and only one of lack of positive parenting (Pianta warmth), indicators of maternal depression, and receipt of means tested benefits. Those retained for the FSP model include whether the child was the first and early motherhood, a range of indicators related to lack of positive parenting (Pianta warmth, home observation, not reading to child, use of childcare, and irregular bedtimes), but also one negative indicator (home observation), maternal health but not depression, worklessness and housing tenure. Although these contrasting indicators are from similar domains it is perhaps surprising that there are so many differences in the predictors of these two outcomes. The measure of poor or fair child health only applies to four percent, so fewer indicators are retained: gender, ethnicity and net household income are also included, as for SDQ and FSP; only irregular mealtimes and bedtimes are retained from among the wide

range of parenting indicators; both maternal health and maternal depression matter here; and lastly receipt of means tested benefits is retained.

In the broadest generalisation delayed childbearing and positive parenting, good maternal health, maternal education and employment status matter for overall development, whilst family context and especially lack of conflictual or negative parenting and lack of maternal depression, as well as some maternal qualifications and higher incomes matter for behaviour. These broad generalisations are retained even when the results are controlled for measures on these outcomes at age 3.

6.2 Behaviour scores

We now turn to a more detailed consideration of the results. The actual categories retained in the final models, the odds ratios and their associated p-values are shown in Table 9 for all three outcomes. In order to emphasise especially strong associations, those significant at the level p<=0.005 are highlighted; where indicators are included in the model that are not significant at the level p<=0.05 (such as controls for additional categories to maintain a particular reference group) these are shown in parentheses.

Figure 2 presents the odds ratios relating to the retained factors for the indicator of very poor behaviour, a score of 14 or higher (up to 34) on the SDQ (strengths and difficulties questionnaire), being the worst 8.5 per cent of such scores; the full bar shows the odds ratio net of the other retained factors, but without controls for SDQ, cognitive scores and health status outcomes at age 3, whilst the smaller bar shows the net odds ratios after such controls.

The most striking result is the very powerful association of poor behaviour outcomes at age 5 with a high score on the Pianta conflict scale: without control for SDQ scores at age 3, those with the highest Pianta conflict scores (27-35, just over four per cent) have a net odds ratio of having extreme SDQ scores at age 5 of over 8:1; those with scores of 21-26 (about 15 per cent) have net odds ratio of 3.6:1; and those with only a moderate indication of parent child conflict (score 16-20, around 25 per cent) have an odds ratio of 1.75:1; all are compared with the low conflict group with scores of 7-15 on the Pianta conflict scale (about 50%). Moreover, these associations with poor behaviour at age 5 remain quite powerful even after controls for SDQ scores at age 3: the odds ratios are reduced to 3.8, 2.0 and a statistically insignificant 1.3 for the three groups ranked by decreasing levels of conflict.

Further indicators relating to negative aspects of parenting are also retained in the model that does not include the age 3 outcomes as predictors: shouting at the child

daily if naughty (OR=1.5); irregular mealtimes (OR=1.7); frequent disagreements with partner over the child (OR=1.6); and, a slightly more tenuous link, whether anyone smoked in the same room as the child (OR=1.4); in addition, the lowest score for the Pianta warmth scale links to poor behaviour (OR=1.7). Taken together with the very strong associations with the Pianta conflict scale, these provide a powerful indication that positive parenting matters and that conflict in the home with the child or the partner has negative impacts on poor behavioural outcomes. Since many of these differing negative parenting will overlap, the combined impact of negative parenting is undoubtedly large; however, they do all add independent explanatory power to the model, despite such overlaps.

Children of depressed mothers experience greater incidence of behaviour problems (OR=1.59 for Kessler score of 7 or higher, and also OR=1.51 for Kessler score of 4-6); in addition, if the mother was currently being treated for depression (OR=1.57), this further reinforces the impact of depression, such that children of mothers with a Kessler score of 7 or higher who are also being treated for depression have a combined OR of 2.50 (=1.59*1.57) for poor behaviour.

There are also strong associations of very poor child behaviour with income (or to some extent indications that children being raised in households in the top income quintile are at much lower risk of behaviour problems). Compare with this top quintile, those in the next highest income quintile band have a OR of 1.82 of poor behaviour outcomes, whilst those in the bottom 60 per cent only have a marginally greater increased risk of poor behaviour (OR=2.20) with there being little indication (not statistically significant) of there being differences among the three lowest income quintile groups. However, the risk of poor behaviour outcomes is further increased when the household is in receipt of non-universal benefits (OR=1.56); since virtually all such households will be among those with the lowest 60% of incomes, the combined effect would be an OR of 3.45.

There is also powerful evidence (p<0.005) in the model without outcomes at age 3 that extreme behaviour problems occur for: boys (OR=1.8); ethnic minorities, both those of Pakistan or Bangladesh origins (OR=2.1) and other non-white groups combined (OR=1.8); and children of mothers without any qualifications (OR=2.0). In addition there is weak evidence that intermediate levels of qualifications (NVQ levels 1, 2, and 3) are also linked to poorer child behaviours – an overall Wald test shows that these three groups do differ significantly from the reference group (NVQ levels 4 & 5), though not reaching significance individually, even though levels 2 and 3 were combined.

Finally, there is some indication of an association of poor child behaviours with partnership context at age 3; surprisingly, in the light of the bivariate associations, only children whose carers cohabit show a significant association (OR=1.42, p=0.008, compared with the bivariate OR of 2.01) with children living with a lone parent being marginally insignificant (OR=1.33, p=0.052, compared with the bivariate OR of 3.43). It seems likely that other characteristics associated with lone parenthood, such as depression, non-universal benefit receipt, or parenting stresses, may be reducing the association with lone parenthood more than for cohabiting parents.

Once further controls for the outcome measures at age 3 are included most of the associations discussed above are attenuated somewhat. Unsurprisingly, the SDQ score at age 3 is a hugely powerful predictor of a very high SDQ score at age 5: those with an age 3 score of 12-15 have an OR of 2.79 and those in the extreme group at age 3 have an OR of 8.07 of having an extreme SDQ score at age 5. In addition, children who performed very poorly on the cognitive tests at age 3 are independently more likely to have behaviour problems at age 5 (ORs=3.48 and 1.44 for Bracken scores of 50-69 and 70-84 respectively), as are children with limiting longstanding illness at age 3 (OR=2.03). Taken together, these strong associations of outcomes at age 3 with poor behaviour at age 5 are bound to account for some of the pathways identified in the model that excludes these age 3 outcomes. What is perhaps surprising is that so many associations with the other indicators are retained in this model. The only age 3 indicators that are removed, or become insignificant though retained for other reasons, are very low Pianta warmth scores, shouting at the child, smoking in the same room, and intermediate qualification levels. In addition the second income quintile and a Kessler depression score of 4-6 become marginally not significant, although combining the two Kessler categories, which have very similar odds ratios would retain the 4-6 score group: a combined Wald test indicates this.

6.3 Foundation Stage Profile

The results for poor learning and development outcomes, as measured by those in the lowest decile of scores on the Foundation Stage Profile, are shown in Table 9 and in Figure 3. Once again we consider the results for the model that excludes the outcomes at age 3 initially and then look at the implications of adding these.

The most powerful association is that with the age of the child, indicated by the grouped dates of birth, since the reports are all from the same time in the school year. The youngest children, those born in July or August 2001, show much higher risk of poor FSP scores (OR=5.00); those born in April to June 2001 have an OR of 2.7 and those born in January to M arch 2001 an OR of 1.86; the reference group

comprises the children born in September to December 2000. It is thus essential to control for child's age in this analysis[1] (though not those for the other outcomes).

There are then an important group of measures related to parenting, although these emphasise the lack of positive parenting rather than the negative parenting aspects that came through so strongly for behaviour problems at age. Poor learning and development outcomes are strongly associated with: very low scores on the positive aspects of home observation (largely a failure to interact positively with the child, showing fewer than 3 of a possible six) with an OR of 2.40; furthermore any negative home observation (OR=1.43); a low score on the Pianta warmth scale (7-29) with an OR of 1.86; reading to the child less than daily (OR=1.24); irregular bedtimes (OR=1.74); and those who had never used child care or nursery before school, perhaps hindering social development (OR=1.41).

Low FSP scores are also strongly related to the mother's educational achievement: compared with those with NVQ levels of 3 or higher, those at NVQ level 2 have an OR of 1.38, those at level 1 an OR of 1.90 and those with no qualifications an OR of 1.63. This last value clearly merits comment, since it might be expected to be higher than that for those at level 1; the raw odds ratios indicate greater risk of a very low FSP score for the no qualification group (5.83) than the level 1 group (4.15); the greater association of lack of qualifications (compared even with level 1) with several of the other indicators (possibly including early childbearing, Pakistan/ Bangladesh origin, and mother's health) probably underlies this apparent paradox.

Children born to mothers who had their first birth whilst young generally experience poorer learning and development outcomes: for a teenage first birth the OR is 1.54 and for one at 20-24 the OR is 1.31. In contrast, first born children are less likely to have low FSP scores (OR=0.77, or equivalently second and higher order births have a developmental disadvantage with OR=1.30).

The mother's general health (but not specifically depression, in contrast to the SDQ analyses) is related to poor FSP scores: where it is reported as good (contrasted with excellent) the OR is 1.25 (marginally significant) and where it is fair or poor the OR is 1.52.

There are a range of other quite strong associations with very low FSP scores: boys develop more slowly (OR=1.92); children living in workless households at age 3 (OR=1.72); those not living in owner-occupied housing (OR=1.31); those in

---

[1] Whether such substantial variations in development by age are plausible and their ramifications for how teachers interact with the child subsequently when the score is on record are considerations beyond the scope of this report.

households with modest to low incomes (the lowest 60%) with an OR of 1.45; and children of Pakistan or Bangladesh ethnic origin (OR=1.64).

When the child outcomes are added we again see very strong associations of these with a poor FSP score. Unsurprisingly it is now the cognitive development score (Bracken school readiness) at age 3 that is most powerfully associated with learning and development scores at age 5: the small group with scores below 70, very delayed readiness, have an OR of 5.78; the somewhat delayed group (scores of 70-84) also have a very high OR at 3.35; in contrast the advantaged group, with scores of over 115 have an OR of 0.428:1 (or odds of 2.31 on); all are compared with the average readiness group with scores of 85-115. However, we also see residual predictive power for the other age 3 poor outcomes: behavioural problems as indicated at age 3 by an SDQ of 12-15 or 16-30 give ORs of 1.36 and 1.72 respectively; and poor child health, indicated by a reported longstanding illness with or without limiting of activities give ORs of 1.36 and 1.72 respectively.

Despite these clear and strong associations with outcomes at age 3 several other indicators remain powerful independent indicators of poor learning and development scores: gender and age; mother's age at first birth before age 25; a cluster of parenting measures – low Pianta warmth, not reading to the child daily, non-use of child care prior to school, and irregular bedtimes; poor or fair maternal health; mother's NVQ levels; and household worklessness. Those that drop out include ethnicity, birth order, home observation scores (both positive and negative), lowish income, and housing tenure.

6.4 Child Health

Poor or fair general child health only applies to four per cent of the sample. Since this outcome is rarer than very high SDQ or very low FSP, fewer factors are retained in the final models; moreover child health has a more direct biological basis and these aspects may not be well captured in the MCS. The results are shown in Table 9 and in chart form in Figure 4.

The most powerful associations are with the mother's own health, both her self-reported health and her depression score. Since it is the mother who also reports on the child's health it is possible that this link reflects a reporting bias, but it is also plausible that there are intergenerational associations in health status. If the mother's health was good (as opposed to excellent) poor child health has an OR of 1.49; if her own health was poor or fair, the OR for poor or fair child health is 2.96; moreover, if there is any indication of maternal depression, judged by a Kessler score of four or higher, child health is again compromised (OR=1.52).

34

Children in moderate to low income households, the lowest 60%, have greater risk of poor or fair health (OR=2.01); this is reinforced where there is receipt of non-universal benefits (OR=1.54); in combination these suggest an OR of 3.12 for those in receipt of benefits since virtually all are also lower income households. Poor or fair child health is also associated with: both irregular mealtimes (OR=1.76) and bedtimes (OR=1.36); Pakistan or Bangladesh ethnicity of the child (OR=1.61); and being male (OR=1.38).

When the outcomes at age 3 are introduced we again see, reassuringly, that the indicators of child health at age 3 are the most powerful predictors of child health at age 5: if the child had a non-limiting longstanding illness at age 3 the OR for poor or fair health at age 5 is 2.23; if the longstanding illness was limiting the OR is 6.56. Behaviour problems at age 3 also predict poor or fair child health at age5, net of health status at age 3: an SDQ score of 12-15 has an OR of 1.74, whilst the poorest behaviour, a score of 16-30 has an OR of 2.14. However there is no indication of an independent link from cognitive performance at age 3, as measured by Bracken school readiness, to health outcomes at age 5.

Once again, despite such strong associations with the outcome measures at age 3, we see that several of the other indicators retain an independent association with poor or fair child health at age 5: gender, ethnicity, irregular mealtimes, mother's health, income and benefit receipt; moreover the strength of these associations remains almost the same as in the model without the age 3 outcome measures. However, the weaker link to irregular bedtimes and the one to maternal depression are not retained.

6.5 Summary reflections

Perhaps the most striking results from these analyses that explore the age 3 measures as predictors of outcomes at age 5 is that so many associations persist even after the introduction of very powerful controls for similar outcomes at age 3. There is some attenuation of odds ratios, but usually not that much, and a few factors are no longer retained. However, many relationships persist, which suggests that these factors play a part in the trajectories to problems for behaviour, development and health between ages 3 and 5.

6.6 Overall Predictive Power of the Age 3 Factors

Since the goal of the PreView tool is to obtain as much predictive power as can reasonably be gained it is important to explore how well our selected models do in

these terms. As discussed in section 3.3 we view the best summary indicator of the predictive power of our models as being the ROC statistic, further described below.

Table 9 includes the summary statistics that were discussed in the sub-section on judging predictive power of the models in section 3.3. As discussed there, the overall fit measures provided by the F-statistic and the Wald chi-square statistic are not very useful for making comparisons between the models, although they are reported here for reference purposes. More useful for comparing the overall fit of the models is the BIC statistic (see above), where a lower value suggests a better overall fit and model complexity is penalised so as to avoid illusory improvements in fit from adding many further parameters. Comparisons across non-nested models, as here, are claimed to be more valid than for the other goodness of fit measures presented (Raftery 1995).

Figure 5A shows the ROC curve (see methods section above) for the two models considered above for the SDQ; comparisons of non-nested models are deemed appropriate. A model that resulted in a perfect classification would achieve 100% sensitivity (identifying all cases correctly) and 100% specificity (identifying all non-cases correctly or no false negatives) and lead to an area under the ROC curve of 1.0. A model that had no predictive power would fall on the reference line. The area under the ROC curve can thus be interpreted as indicative of the overall effectiveness of the model in discriminating cases (those with high behaviour problems at age 5) from non-cases (those not classified as having such behaviour problems at age 5. For the model that omitted the outcomes at age 3 the area under the ROC curve is 0.816: this can be interpreted as the probability that the model will rank a randomly chosen child with serious behaviour problems at age 5, as measured by the SDQ, as being more likely than a randomly chosen child without such problems at age 5 to have such behaviour problems; it is thus a summary measure of the classification power of the model. When the outcomes at age 3 are also included in the model the area under the ROC curve rises to 0.853, showing an unsurprising improvement in classificatory power; the difference in areas under the curve is highly statistically significant (p<0.00005) with non-overlapping 95% confidence intervals (not corrected for survey design).

Figure 5B shows the ROC curves for the two models for low FSP scores, with the area being 0.802 for the model omitting age 3 outcomes and 0.826 when these are included; again this difference is highly significant (p<0.00005), despite being apparently small. Figure 5C shows the ROC curves for poor or fair child health. Unsurprisingly the areas under the two ROC curves are lower than for SDQ or FSP, given the more extreme nature of the outcome, the likely role of biology and fewer

significant predictors: for the model without age 3 outcomes the value is 0.725 and for that including the outcomes 0.76; again these differ significantly (p<0.00005).

7. What extra explanatory power can be derived from MCS1 measures?

The earlier report (Kiernan and Mensah 2009) examined the predictive power of measures available around or before birth, but used several measures at age 9 months when the first wave of MCS data was collected. As discussed earlier, we have brought a few other key measures into this analysis, which results in slightly different selected models based on the age 9 months data; specifically we have included: birth weight, which was treated separately in the previous report ; breastfeeding patterns; whether the mother was overweight or obese prior to the birth (this measure refers to pre birth characteristics, but was not available at the time of the previous analysis); and the mother's attitudes to parenting required for successfully bringing up children on several dimensions, such as cuddling, talking, establishing regular patterns.

In Figure 6 we summarise the indicators that are retained in the revised model that only uses information from the age 9 month study for each of the child outcomes at age 5. Those that survive into a 'final' model that only retains those factors that are most powerful predictors of the outcomes at age 5 after consideration of all indicators and outcomes measured at age 3 as well as those from birth and age 9 months are highlighted in Figure 6. Several of the indicators that are not retained in the final model are replaced by a more powerful association with a similar measure at age 3, such as socioeconomic status, partnership status, and mother's health or depression; however in a few instances, such as mother's health in relation to the child health outcome both measures at age 9 months and at age 3 are retained. These details are discussed more fully subsequently.

As for the analyses using age 3 information, we see that several of the 'fixed' characteristics that can be equally well measured around birth or age 3 matter: boys have worse outcomes on all three measures at age 5; ethnicity has a lasting influence on behaviour and health outcomes and some for development. Perhaps more importantly we see that low birth weight and breastfeeding duration are both associated with worse outcomes for all three age 5 measures and these are powerful enough for FSP and health to persist even after controls for age three indicators; it is not certain that birth weight or more particularly duration of breastfeeding will be well measured by retrospective reports when the child is age 3. Birth weight data may well be available from maternity records.   The mother showing poor attitudes to child development is associated with worse outcomes both for behaviour and development at age 5, though this disappears after controls for parenting measures at age 3. Although less clear than at age 3, partly because measures are less refined, there are again indications that negative aspects of parenting play through more for

poor child behaviour than for poor child development (e.g. mothers who do not give up smoking during pregnancy and who were not very happy about the pregnancy, in addition to showing poor attitudes and cursory or no breastfeeding). Very late or no entry to antenatal care is related to poor health outcomes at age 5.

Mother's health and depression at age 9 months matter for all three outcomes at age 5; her self-efficacy being low, perhaps linked also to weight problems during pregnancy, only shows a clear association with poor child behaviour at age 5 and not to development or child health. A range of socioeconomic measures, including mother's education, housing tenure and IMD relate both to poor behaviour and poor development; low income is associated with poor development and child health; and children in workless households at age 9 months have poorer health outcomes at age 5.

7.1 Behaviour scores

Table 10 shows the selected predictive factors for the revised model around birth/ age 9 months, the two models examined earlier for age three year predictors (excluding and including outcome measure at age 3) for reference, and the model that retains the best predictors from both around birth/ 9 months and age 3 including outcomes measured at age 3. In discussion we shall concentrate on the first and last of these models since the ones for age three factors have already been presented at length in the earlier sections. To aid in identifying the key results those measures at birth/ age 9 months that are retained in the final model are highlighted in the left hand column (mother's weight, smoking through pregnancy, high malaise score, and lack of self-efficacy); moreover those measures that have very high levels of statistical significance (p<=0.005) are highlighted in the body of the table for the relevant models.

As discussed earlier, there are a range of 'fixed' characteristics that can be measured at birth, (9 months), or age 3 and several of these are retained in the overall final model: gender of cohort member and whether a first birth and ethnicity. There are other characteristics that can change over time, but probably not by very much, including mother's NVQ level: the inclusion of the age 3 NVQ levels rather than the age 9 month ones is a very marginal choice and either would suffice. The measures of socioeconomic status probably do change a little more rapidly and it seems that the more recent (or proximate to age5) measures distinguish behaviour scores more sharply.

Yet other measures are available but not fixed over time, where the impact and interpretation may be different. There are two instances of such measures stand out

in relation to child behaviour scores. The mother smoking through pregnancy matters far more than the measure of whether someone in the household smoked in the same room as the cohort member at age 3; the strong pressures not to smoke during pregnancy probably mean that such a measure indicates more than smoking and partially captures 'responsible' parenthood as well as possible iatrogenic influences on the foetus (e.g. Pickett et al 2008). The second measure of this class relates to maternal depression: once again it is the measure at age 9 months, which proxies for post-partum depression, rather than the indicators at age 3, that matter most for behaviour outcomes when the child is age 5. We note that this comparison is made less conclusive by the fact that the measures used at age 9 months and age 3 were different.

Finally there are a range of predictive factors that were measured either at birth/ age 9 months or at age3, which sometimes might have been replicable at the other age but are not available here. Thus the measure of mother's self-efficacy occurred at age 9 months but could in principle have been measured at age 3 and the timing of such measurement may matter but likely does not. Most of the parenting and parent-child interaction measures we have used come from the interview at age 3 and were not available for the age 9 months survey, although some of these might have been collected (e.g. smacking, shouting, partner disagreement over child,)[2]. Many of the measures drawn from the age 3 interview would not be replicable as young as age 9 months, including both the SDQ and Bracken scores and the Pianta scales.

7.2 Foundation Stage Profiles

Table 11 shows the results for the models for Foundation Stage Profile scores at age 5, with the same materials as presented in Table 10 for the SDQ scores. The birth/ 9 month characteristics that show persistent leverage even after control for age 3 characteristics and outcomes are: gender, age and ethnicity, all fixed characteristics that can equally well be measured at age 3; birth weight a fixed characteristic, but better measured at birth to avoid recall bias; duration of breastfeeding, again better measured at age 9 months than retrospectively at age 3; whether there were three or more children in family home (probably somewhat similar at age 3); housing tenure (generally slow changes, except where precipitated by partnership breakdown for example) and where housing tenure at age 3 was a close competitor; and the index of multiple deprivation (which had identical values at age 9 months and age 3 except for migrants, since the same reference year was used).

---

[2] There are some alternative measures of parenting and child temperament that were collected at age 9 months which we have not been able to incorporate here

As for the SDQ, mother's NVQ levels at age 3 marginally outperformed the same measure at age 9 months, as did the mother's reported general health at age 3 compared with 9 months; for both measures a single report at either age would suffice. The socioeconomic status indicators include both tenure at 9 months and the IMD at 9 months (for both an observation at age 3 would also suffice) and whether the household was workless at age 3. Otherwise, as for the SDQ, the parenting measures all come from the interview at age 3, as do the intermediate outcome measures: measurement is likely to be much more reliable or possible for these characteristics at age 3 than at 9 months. We not that the differences between SDQ and FSP in terms of which parenting variables matter is unchanged from the earlier discussion, with conflictual/ negative measures linking to poor behaviour outcomes and lack of positive parenting/ warmth linking to the FSP.

7.3 Child Health

Table 12 shows the parallel results for the cohort member's health being poor or fair at age 5. Once again gender and ethnicity matter considerably; both low birth weight and lack of breastfeeding also have persistent effects even after control for age 3 characteristics and outcomes; we again note the preference for measuring these at birth/ 9 months. A further fixed characteristic that just happens to be measured at age 9 months is whether the mother's parents had separated before she reached age 17, which has a persistent link to low child health (although the pathway or aetiology is unclear).

The socioeconomic predictors of low child health are dominated by household income and, once again, the measure at age 3 shows slightly stronger associations although substitution of age 9 month measures would not lose much predictive power. Interestingly, we see that *both* measures of the mother's general health at age 9 months and at age 3 are retained in the most complete model: if the mother's health was poor or fair at either point it raises the chances of the child's health being poor or fair at age 5; moreover this is a time varying characteristic where having both measures add leverage to prediction (over and above knowing whether the child had a limiting or non-limiting longstanding illness at age 3). We cannot be certain whether this compounded and persistent association of mother's self-reported general health with the child's general health as reported by the mother represents genuine and plausible intergenerational pathways or, at least partly, represents reporting biases of the mother.

8. Predictors of any poor outcome

As a follow up to the supplement to the earlier report we were asked to look at the predictors of children experiencing any of the three poor outcomes at age 5. Examining multiple poor outcomes proves difficult because the combinations are quite rare. One challenge is that poor behaviour scores on the SDQ and poor development scores on the FSP are over twice as prevalent as low child health. A further challenge is that we have limited the analysis to those cohort members for whom we have information on all three outcomes at age 5, which restricts us to a sample size of 5,616 of whom 17% experience one or more poor outcomes. Essentially, as we would anticipate, the combined model for any poor outcome effectively averages the three separate outcome specific models and more or less gives twice the weight to the ones for SDQ and FSP as to the child health ones.

The earlier analyses showed that there were some quite clear and informative distinctions between the sets of predictors for each of the outcomes, although of course with some overlap. In particular in the final models combining all measures, interesting differences in the types of parenting that are associated with poor SDQ and FSP outcomes at age 5 emerged; the age of the child was hugely important for the FSP but not at all for the SDQ or child health; mother's health at both ages matters for child health, at age 3 only for the FSP, and not for the SDQ; however maternal depression matters for the SDQ.

For these reasons we feel the models for experiencing any one or more of the three poor outcomes blur the picture, but others may well disagree. Since we were explicitly asked to examine this combined outcome measure we present the results in Table 13 which is of the same form as Tables 9-11 for the individual outcomes. The results are sometimes puzzling. The parenting measures, which proved such strong predictors of poor SDQ and FSP scores almost disappear: the fairly clear cut roles of conflictual parenting for the SDQ and lack of warm parenting for the FSP are largely lost in the blurring of the outcomes. The combination of 'apples and pears' also seems to lead to a wider range of socioeconomic measures being retained: housing tenure and IMD at age 9 months and income and benefit receipt at age 3. Both poor/fair maternal health at age 3 and the proxy for post-natal depression from age 9 months are retained, as is an alternative indicator of maternal health status at 9 months, having a non-limiting longstanding illness (but not a limiting one).

9. **Summary and Conclusion**

The main focus of this report has been on the analysis of the information collected at age 3 and how well this predicts three key outcomes at age 5, poor behaviour scores on the Strengths and Difficulties Questionnaire, poor developmental outcomes on the school Foundation Stage Profile, and poor or fair reported child health. A very wide range of factors were considered and strategies were implemented to remove many indicators that did not retain significant predictive power when other factors were included. The early sections of the report show the associations with each of the three outcomes at age 5 for all of the measures considered without any control for other measures. Once other factors were controlled many of the measures gave no leverage in predicting any of the outcomes: these were removed from the models to avoid unnecessary clutter and complexity.

9.1 Changing Family Experiences from Infancy to Age 3

Section 5 described relationships for the three age 5 outcomes with combinations of measures on similar measures of status of the family at age 9 months and again at age 3. The purpose was to illustrate the issues raised by the contrast of the family being persistently disadvantaged at both waves, experiencing episodic disadvantage at one wave but not both and not experiencing such disadvantage. We were also interested in getting insights into whether the timing of disadvantage for the episodic group, or alternatively, whether improvement or deterioration of status mattered more. The measures examined included poverty status, employment status, maternal depression, and maternal general health; in addition, we examined associations with partnership breakdown of the child's parents between age 9 months and age 3.

In general these descriptive results show that children in families that experience persistent disadvantage show the greatest risk of poor outcomes, but that episodic disadvantage is also associated with poorer outcomes, with typical odds ratios being around 4:1. There is little clear evidence of timing effects for the episodic experience of family disadvantage, with typical odds ratios being around 2.5:1 for either, although there are suggestive indications that more recent experience (moving into disadvantage) of unemployment and poor maternal health may matter slightly more for the child's development and behaviour. It is also worth noting that the persistently disadvantaged groups have odds ratios that are the below the product of the two episodic disadvantage ones (e.g. 2.5*2.5=6.25>4), showing that the reinforcing effect of persistent disadvantage is less than entering disadvantage newly.

9.2 Commonalities in retained predictors across outcomes at age 5

In the detailed main analyses (section 6) we show that measures from a wide range of domains retain predictive power: child and family (demographic) characteristics, parenting and parent-child interplays, maternal health (both mental and general), maternal qualifications, and socioeconomic situation. Thus there is no single 'magic bullet' domain for policy and practice interventions: identification of children who are likely to have poor outcomes requires information from all of these domains. However, we have been able to reduce the range of information collected in the Millennium Cohort Study at age 3 to more manageable proportions through retaining only those measures that matter for the three outcomes. In order to be able to implement monitoring at age 3 into PREview there will have to be informed judgements about how to further reduce some of these key predictors into those that are able to be collected simply within the time constraints. However, such further simplification was not part of our brief.

One of the remarkable findings from our analyses is that many of the predictive factors are still powerfully related to each of the outcomes even after we control for similar measures on all three outcomes at age 3. A few lose significance but many retain very similar odds ratios even after control for the outcomes.

Whilst we retain predictors from each of the main domains (demographic, parenting, health, qualification, and socioeconomic situation) for each outcome it is noteworthy that the details are often different (see Figure 1 and Table 9). The only indicators that are retained for all three outcomes at age 5 are gender, ethnicity, and moderately low household income: boys have significantly poorer outcomes than girls; those of Pakistan or Bangladesh origins show consistently poorer outcomes than others; and children living in households in the lowest 60% of income band show consistently poorer outcomes. Further indications of the role of poverty are provided by inclusion of either receipt of means tested benefits (for SDQ and health) or household worklessness (for FSP), which overlap substantially, but might also be capturing slightly different aspects at the margins. Similarly there is overlap, both conceptually and in incidence, for the indicators of irregular bedtimes (FSP and health) and irregular mealtimes (SDQ and health).

9.3 Differences in retained predictors by outcome at age 5

Although the different domains are all represented with retained predictors for each of the three outcomes at age 5, many interesting and informative differences do emerge. The age (date of birth) of the child matters for learning and development

(FSP), but not for SDQ or health outcomes. For the other demographic characteristics, partnership status of the mother at age 3 is retained for the SDQ whilst early motherhood and first-born children are disadvantages for the FSP.

The sharpest and most interesting differentiation appears for the parenting measures. The SDQ is largely influenced by the negative aspects of parenting (especially conflict as measured by the Pianta scale, but also partner disagreements, irregular mealtimes and shouting at the child, and more tenuously smoking behaviour) with a small residual impact of lack of warmth. In contrast the retained parenting indicators for learning and development, the FSP score, emphatically emphasise the lack of positive parenting (low Pianta warmth score, positive home observation, reading to the child, and possibly irregular bedtimes and use of child care prior to school) with only negative home observation from the more conflictual measures. Perhaps unsurprisingly, given the lower incidence of poor or fair child health outcomes and the biological pathways also involved, fewer parenting measures are retained for this outcome: both irregular mealtimes and irregular bedtimes are associated with poorer outcomes.

Although the mother's health matters for each of the three outcomes at age 5, we once again see different pathways implicated. Maternal depression has a greater impact on behaviour scores, whilst her general health matters more for the child's learning and development. Given likely intergenerational health linkages, it is reasonable that both aspects of maternal health matter for low child health, although it is the mother's general health that is retained after controls for outcomes measured at age 3.

Mother's qualification levels are retained as predictors for both high SDQ and low FSP, but not low child health. However, it is predominantly lack of qualifications that matters for the child's behavioural outcomes, whilst there is a stronger gradient with mother's qualifications for the learning and development outcome, with strongly significant increases in the incidence of poor learning and development outcomes for children of mothers below NVQ level 3.

In the socioeconomic situation domain we see indications that behaviour problems, a very high SDQ score, are much rarer for children living in households in the top 20% of incomes, with a small differential for the next quintile (20-40%); among the lowest 60% of household equivalized incomes there is a clear further indication that those in household receiving non-universal benefits have higher incidence of poor behaviour. Poor or fair child health is also significantly more prevalent in household in the lowest 60% for equivalized income and this is again reinforced by being in a household receiving non-universal benefits. It is striking that differentiation on

income begins at such a high threshold for both of these child outcomes. The most powerful retained predictor from this socioeconomic situation domain for poor learning and development, as measured by a very low FSP score, is worklessness in the household; there is some evidence that this is reinforced both for those households in the lowest 60% for income and those that are not owner-occupied. For all three child outcomes at age 5 there is no evidence to discriminate further on income within the lowest 60% of incomes, although benefit receipt and worklessness clearly capture aspects of even lower incomes.

9.4 Outcomes at age 3

Incorporating measures of similar outcomes at age 3 to those measured at age 5 inevitably adds predictive power, but perhaps surprisingly only moderates the associations with many of the measures considered above.

The SDQ measure at age 3 is of the same indicator as that used for age 5, although approximately the two lowest deciles are distinguished. A very poor behaviour score at age 3, an SDQ score of 16-30, is a very powerful predictor of a very poor behaviour score at age 5 (OR=8.1). A moderately poor behaviour score at age 3, an SDQ of 12-15, is also quite powerfully predictive of a very poor behaviour score at age 5 (OR=2.8). Furthermore, children who have delayed development at age 3, as measured by a Bracken school readiness score of below 85, and those who have a limiting longstanding illness at age 3 both have a higher prevalence of very poor behaviour scores at age 5, being especially large for the small fraction with very delayed development.

The Bracken school readiness indicator at age 3 is a measure of development and unsurprisingly is a powerful predictor of poor learning and development at age 5, indicated by a very low FSP score. However, we once again see that poor indications on the other age 3 outcomes, behaviour and child health, serve to reinforce the prospects for poor FSP scores.

The indicator of child health at age 3 distinguishes those with limiting and non-limiting longstanding illness. Once again, poor child health at age 3 is a hugely powerful predictor of poor or fair child health at age 5, particularly if the illness is limiting (OR=6.6). Poor or fair child health at age 5 is also considerably more likely for those with poor behaviour scores at age 3; however there is no residual association of child health at age 5 with the Bracken scores at age 3.

9.5 Persistent Predictors

Given the hugely powerful associations of each of the child outcomes at age 5 with the parallel outcome at age 3 and the quite strong reinforcement from knowledge of the other outcomes at age 3 it is remarkable how powerful many of the other retained predictive measures are after such controls. Relatively few are removed even by the fairly simple criterion of moving to a p-value of above 0.05. Even more surprisingly many retain very similar odds ratios and become only marginally less statistically significant. For the SDQ at age 5 there is very little reduction in the predictive nature of gender, ethnicity, several parenting measures, maternal mental health or socioeconomic situation. The largest attenuation for the SDQ retained predictors is for the Pianta conflict score, where the odds ratio for the extreme category is reduced from 8.2 to a still huge 3.8. For the FSP at age 5 there is very little reduction in the predictive nature of gender, age, early motherhood, most of the parenting measures, poor maternal health, maternal qualifications and household worklessness. For poor or fair child health at age 5 there is little change in the predictive nature of gender, ethnicity, irregular mealtimes, maternal general health, or socioeconomic situation.

9.6 Predictive performance

The predictive performance of the models is assessed using the Receiver Operating Characteristic (ROC) measure. This provides an indication of how well the models perform in terms of the probability that a random pair of one child with the poor outcome and one without would be correctly ranked by the predicted probabilities from the model. The models for the SDQ and for FSP perform similarly, with about 80% predictive power for the models that do not include the outcomes at age 3 and rising to nearer 85% when these are included. Given the lower prevalence of poor or fair child health and its clearer biological origins, it is to have been expected that predictive power for this outcome would be lower: 72.5% without age 3 outcomes and 76% when they are added.

9.7 Addition of birth/ age 9 month information

A further key question was to explore the extent to which measurements from the first wave of MCS, pertaining to birth and up to age 9 months, retain predictive power in addition or substitution to the measurements at age 3. Three attributes that were not included in the previous report, duration of breastfeeding, whether the mother was overweight or obese during the pregnancy, and the mother's attitudes to child development were considered as additional predictors of the child outcomes at age 5, plus birth weight which was considered as an add-on in the

previous report. Since these additional measures are introduced the first step was to re-estimate the models based only on the birth/ age 9 month measures. We could then ask which of these measures should be retained in an overall model that combines measures at birth/ 9 months, age 3 and age 3 outcomes.

Several of the key demographic characteristics considered for the models using information at age 3, such as gender, birth date of child, ethnicity of child, whether first born, mother's age at first birth or whether the mother's parents separated before she reached age 17, are fixed over time and thus also part of any model for birth/ age 9 months. In addition birth weight, duration of breastfeeding and whether the mother was overweight or obese during pregnancy are fixed forwards, but either better measured around the time of the birth to avoid recall bias (birth weight and breastfeeding) or possibly simply a single observation of a characteristic that can vary over time (e.g. mother's BMI). Many other measures, of partnership status, parenting, mother's health and well-being, and socioeconomic situation are time-varying and can generally be better measured contemporaneously.

Once again we see (Figure 6) that measures from each of the broad domains at birth/ age 9 months (demographic/ mother's characteristics, parenting measures, maternal health and well-being, and socioeconomic situation) are retained as having predictive power for each child outcome at age 5. In broad terms we see that both low birth weight and short or no breastfeeding are associated with poorer outcomes at age 5 for all three measures of behaviour, development, and health. As with the models using the information at age 3 we see indications that negative aspects of parenting (smoking through pregnancy, being unhappy about the pregnancy) are clearly linked to poor behaviour, whereas both breastfeeding and poor attitudes to child development link to poorer outcomes on learning and development as well as child behaviour. We again see that mother's mental health matters more for child behaviour and her general health more for learning and development, although the distinction is less clear than for age 3 measures.

9.8 Additional leverage from birth/ age 9 month information

Of most interest here are those indicators from birth/ 9 months, excluding the fixed characteristics, that retain predictive power over and above the information from the survey at age 3.

For predicting poor behaviour at age 5, as measured by a high SDQ score, knowledge that the mother was overweight or obese during the pregnancy, that she smoked through pregnancy, that she experienced post-natal depression and she lacked self-efficacy (not measured at age 3) all added significant leverage to the model. Smoking

during pregnancy and post-natal depression both retain predictive power and give more leverage than parallel measures at age 3.

Poor learning and development at age 5, as measured by a low FSP score, has additional predictive power from knowledge of birth weight, duration of breastfeeding, whether there were three or more children in the home, housing tenure and the area index of multiple deprivation (IMD). Although tenure and IMD were also available from age 3, the measures at 9 months gave marginally more leverage in accounting for low FSP; in contrast similar close run measures of mother's general health, mother's NVQ level and workless households from age 3 marginally outperformed those from age 9 months. In all these cases it would probably be adequate to have either measure.

For predicting poor or fair child health added leverage comes from knowledge of birth weight, lack of breastfeeding and the mother's general health, as well as whether her parents separated before age 17 (which is really a fixed characteristic of the mother). Mother's general health provides an example of where it adds predictive power to the model for poor learning and development outcomes to have observations on this measure both at age 9 months and at age 3.

9.9 Experience of any of the three poor outcomes at age 5

We also present the results of similar modelling exercises for a combined indicator of experience of any one or more of the three poor child outcomes. We remain unconvinced that this combination is useful, as explained in the text, since it results in a significant loss of sharpness of focus for some of the retained predictors that really matter for the individual outcomes, being essentially a weighted average of the three separate models with the SDQ and FSP getting about twice the weight of the health outcome. As such, it probably adds nothing and loses key contrasts, such as the clear distinction between the types of parenting measures that matter for the SDQ (negative/ conflict) and for the FSP (lack of positive or warm parenting).

10. References

Bracken, B.A. (2002) "Bracken School Readiness Assessment Administration Manual" Psychological Corporation

Case A, Lubotsky D & Paxson C. (2002) Socioeconomic status and health in childhood: the origins of the gradient. *American Economic Review* 92 (5), 1308-1334

Dex S & Joshi H (2005) *Children of the 21st century: From birth to nine months* (Bristol, The Policy Press).

Goodman R. (1997) The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.

Goodman R. (2001) Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337-1345.

Hansen K (2008) *Millennium Cohort Study first, second and third surveys: a guide to the datasets (third edition)* (London, Centre for Longitudinal Studies, Institute of Education).

Hansen K & Jones EM (2008) Foundation Stage Profile and devolved administration teacher survey, in Hansen K & Joshi H. *Millennium Cohort Study, Third Survey: A User's Guide to Initial Findings* (Centre for Longitudinal Studies, Institute of Education, London).

Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, Walters EE, Zaslavsky AM (2002) Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychological Medicine 32:959-976

Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, Howes MJ, Normand SL, Manderscheid RW, Walters EE, Zaslavsky AM (2003) Screening for serious mental illness in the general population. Archives General Psychiatry 60:184-189

Ketende S. (2008) *Millennium Cohort Study: Technical Report on Response, 2nd edition* (London: Centre for Longitudinal Studies, Institute of Education).

Kiernan KE & Huerta MC. (2008) Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood, *British Journal of Sociology*, 59(4), 783-806.

Kiernan KE & Mensah FK. (2009) Poverty, maternal depression, family status and children's cognitive and behavioural development in early childhood: a longitudinal study, *Journal of Social Policy*, 38(4): 569-588.

Mensah, F.K and Kiernan K.E (2009) Parents' mental health and children's cognitive and social development. *Social Psychiatry and Psychiatric Epidemiology published on line pre-publication*

Office for National Statistics. (2003). *Ethnic group statistics: A guide for the collection and classification of ethnicity data* (London: HMSO).

Office of the Deputy Prime Minister (2004) *The English Indices of Deprivation (revised).* (London: ODPM)

Pianta RC. Child-Parent Relationship Scale. Charlottesville: University of Virginia, 1995.

Pickett, K E, C Wood, J Adamson, L DeSouza and L S Wakschlag (2008) Meaningful differences in maternal smokingbehaviour during pregnancy: implications for infant behavioural vulnerability. *Journal of Epidemiology and Community Health* 62:318–324.

Plewis I. (2007) Non-response in a birth cohort study: the case of the Millennium Cohort Study, *International Journal of Social Research Methodology*, 10(5), 325-334.

Qualifications and Curriculum Authority. (2003) *Foundation Stage Profile Handbook* (London, Department for Education and Skills).

Raftery, A.E. 1996. Bayesian model selection in social research. In P.V. Marsden (Ed.) *Sociological Methodology* 1996:111-163. Blackwell, Oxford.

StataCorp LP (2007) Predict – Obtain predictions, residuals, etc., after estimation. In *Stata base reference manual, Volume 2, release 10.* (Texas)

Subramanian SV, Ertel K. (2008) Is the use of self-rated health measures to assess health functioning misleading? *International Journal of Epidemiology*, 37:1436–37.

UCLA Academic Technology Services. (2008) *Regression with Stata: Chapter 3 – Regression with Categorical Predictors.* Available online at: http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter3/statareg3.htm

 (accessed 2nd December 2008)

Appendix A

List of variables from the 9 month survey (MCS1) included in the analysis. New variables not included in the first report are highlighted in italics

Child Characteristics
Gender
Age in months
Birth weight
First born or later
Whether Breastfed

Pregnancy
Pregnancy planned or not
Happy to be pregnant
Time of first ante-natal visit
Smoked during pregnancy
Alcohol consumption
Mother's BMI pre pregnancy

Maternal Attributes
Whether she had experienced parental separation during childhood
Whether she had been in care
Mother ever worked
Level of Highest Qualification at 9 months
Ethnicity
Mother's attitude to her child's development

Maternal Health
General Health
Psychological Distress as measured by Malaise score
Whether felt low or sad since the baby was born
Ever-diagnosed with depression
Measure of self efficacy
Limiting longstanding disability

Demographic and household Characteristics
Parent's relationship status at the birth of the child
Mother's age at first birth
Family Status at 9 months
Number of children in the Household
Language usually spoken in the home

Socio-economic characteristics
Number of household members in work
Household Income
Income poverty
Receipt of means tested benefits
Housing Tenure
Housing difficulties
Area Index of Multiple Deprivation