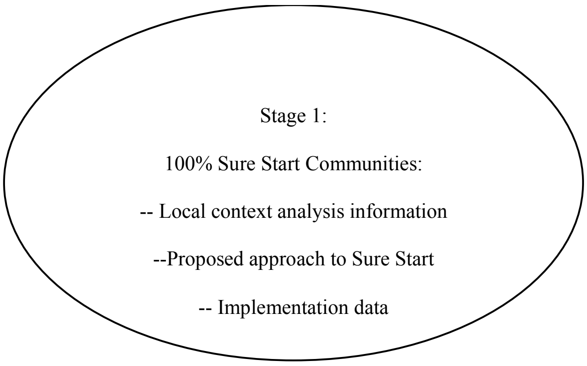
IMPACT STUDY METHODOLOGY REPORT
SEPTEMBER 2002

We have designed a multifaceted investigation to illuminate the effects of Sure Start, as well as the conditions under which and the mechanisms by which it exerts its impact. Central to our approach to impact evaluation is a research design that is strategically staged. The three figures which follow the paragraphs below graphically illustrate the **progressive focussing** nature of our 3-stage approach to impact evaluation. We provide an overview of the 3-stage approach, then discuss issues of control groups, and thereafter further detail the three stages of evaluation for the first 6-year grant period.

I. Overview of Three-stage Strategy

In the first year of measurement for each Sure Start programme (i.e., 2002 for Trailblazers and Wave 2 programmes; 2003 for Wave 3/Wave 4 programmes), data obtained on all Sure Start Programmes as part of the implementation study and the local context analysis will be examined by the impact evaluation team. In particular, we will consider (a) the geographic, demographic, and ethnic nature of Sure Start communities, including their relative levels of deprivation and social exclusion, (b) strategies proposed by each community for implementing Sure Start in their original funding applications, and (c) their early success in implementing Sure Start.

Stage 1:
2001\2002\2003

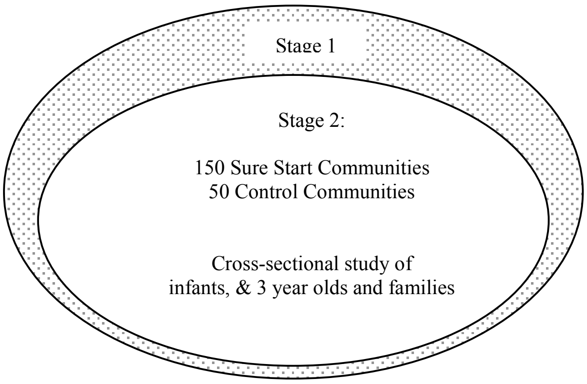


This information will be used to select a subset of 150 Sure Start Programmes for more intensive investigation in 2003 (TB\Wave 2) and 2004 (Wave 3\4) (see figure below). This first round of strategic sampling (which sets the design for Stage 2 of the impact study) will be based upon the following **three** kinds of information, broadly conceived: geographic, demographic, and ethnic characteristics of Sure Start communities, intervention strategies of local Sure Start programmes (i.e. inputs), and information pertaining to the quality of early implementation of Sure Start.

Hence, we plan to base the first-round of strategic sampling on a **combination** of inputs (e.g., demography, intervention strategy) and implementation factors presumed to reflect anticipated outputs.

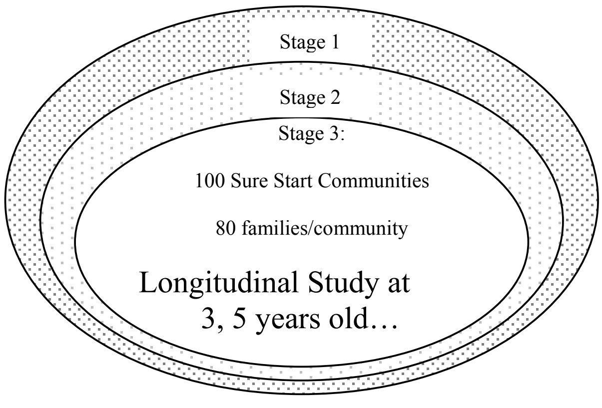
In the second stage of the impact study, we will also randomly select 50 Sure Start-to-be communities to function as control communities. We will collect community level data pertinent to the evaluation of impact in the chosen 150 Sure Start communities and the 50 control communities. For each Sure Start community 100 families will be recruited for study by means of home visits. These 100 families will consist of 80 with a 9-month-old; and 20 with a 3 year old. During these same two years, for each of the 50 control (Sure Start-to-be) communities, 75 families will be recruited for home visits; 25 with a 9-month-old; and 25 with a 3 year old.

Stage 2:
2003\2004



In the third stage of our impact study, some of the Sure Start community families originally participating in the cross-sectional study (Stage 2) will be invited to participate in a long-term longitudinal study. More specifically, the families from each of 100 (of 150) selected Sure Start communities whose infants were studied in Stage 2 (i.e. the cross-sectional study) will be studied, along with their child, when their child is 3 and 5 years of age (and thereafter). In other words, not only will community-level data collections continue in these 100 communities on a longitudinal basis but so, too, will child- and family-level data collections. Thus, a "delayed longitudinal birth cohort" study will characterise the third stage of our strategic research design for investigating the impact of Sure Start. The 100 communities for the longitudinal study will be selected at the second round of strategic sampling.

Stage 3:
2005 and Beyond



The second round of strategic sampling will take place at the end of the cross-sectional data collection (i.e. stage 2). More specifically, we will use evidence on child and family functioning collected in 150 Sure Start communities to identify communities in which children and families are functioning better and worse than would be expected on the basis of background information. Thus, of the 150 communities selected in the first stage of strategic sampling, 100 will be further selected for longitudinal follow up on the basis of **three** sets of factors. Two of these three factors were in the first stage of strategic sampling, but the third is importantly different:

- Geographic, demographic, and ethnic characteristics of Sure Start communities
- Intervention strategies of local Sure Start programmes (i.e., inputs)
- **Cross-sectional data indicating whether children and families in specific Sure Start Communities are being positively affected by Sure Start**

Only after ensuring that we have a good representation of communities on the basis of inputs (i.e. factors [1] and [2] above), will we select communities that systematically vary from excellent to poor in terms of detected effects of Sure Start once differences in population characteristics have been taken into account. This selection on the basis of outputs is made possible by establishing value-added "community effects" within the multilevel modelling approach to analysis. (This is akin to the establishment of school effects in school-effectiveness research and hospital effects in health-services research.) Hence, the second round of strategic sampling is based on a combination of inputs and outputs.

For purposes of comparison, children growing up in economically disadvantaged families in non-Sure Start communities and being studied as part of the Millennium Cohort Study (MCS) being carried out by the Centre for Longitudinal Studies, Institute of Education, with funding from the ESRC, will function as longitudinal controls. More detail is provided below.

Having outlined the major phases of inquiry in our strategically staged impact evaluation, we now provide the schedule of the cross-sectional and longitudinal stages of the study. The overall research design of the impact evaluation is represented in Table 1, and the sampling plan in terms of children/families, which maps on to this design, is shown in Table 2. As can be seen in Table 1, the cross-sectional data collection is conducted in years 2003 and 2004 and the longitudinal data collection, while initiated in these years, are extended through the 2008 funding year. Table 2 stipulates the number of children/families that will be studied in both the cross-sectional and longitudinal phases of the impact study.

Table 1: Research Design Of Impact Evaluation

CALENDAR YEAR

TB/Wave 2	02	03	04	05	06	07	08
9 Months		*CS*/$\mathbf{L_1}$					
3 Years		*CS*		$\mathbf{L_1}$			
5 Years						$\mathbf{L_1}$	
Wave 3/4							
9 Months			*CS*/$\mathbf{L_2}$				
3 Years			*CS*		$\mathbf{L_2}$		
5 Years							$\mathbf{L_2}$
Controls							
9 Months		CS					
3 Years		CS					

CS: Cross-Sectional
L_1: First wave of longitudinal sample – Trail Blazer (TB) and Wave 1
L_2: Second wave of longitudinal sample – Wave 3 and Wave 4

*Note that the fieldwork for the longitudinal study in year 05 through 07 will actually extend 3 months into the following year.

Table 2: Sampling Plan: Number of Children

	CALENDAR YEAR						
TB/Wave 2	02	03	04	05	06	07	08
9 Months		**6k**					
3 Years		1.5k		**4k**			
5 Years						**4k**	
Wave 3/4							
9 Months			**6k**				
3 Years			1.5k		**4k**		
5 Years							**4k**
Controls							
9 Months		1.25k					
3 Years		1.25k					
Total		**10.5k**	**7.5k**	**4k**	**4k**	**4k**	**4k**

Before providing more detail on each of the three stages of impact evaluation, further discussion of the issue of controls/comparisons and the selection of ages of measurement is provided. After providing additional detail on each of the three stages of impact evaluation and the selected ages of measurement, plans related to actual fieldwork are provided.

II. Controls and Comparisons

In an ideal scientific world, the effects of Sure Start would be evaluated using a randomised control design in which economically disadvantaged communities were randomly assigned to receive Sure Start or to continue to receive standard services. Because this strategy was ruled out, it was our initial plan to identify communities that were as economically disadvantaged as Sure Start communities but did not receive Sure Start and treat these as comparison communities. The expansion of Sure Start (from 250 to 500 programmes) undermined the feasibility of this strategy. As a result of these scientifically challenging circumstances, we have developed two approaches to scientific control when it comes to evaluating the effects of Sure Start.

As already noted, the first approach involves the Sure Start-to-be communities. More specifically, these communities which are expected to be more or less comparable in terms of social, ethnic, and economic composition to Sure Start communities, but differ from Sure Start communities in terms of having a Sure Start programme when assessments are carried out, will afford the opportunity--very early in the research process--to make comparisons at the level of individuals, families, and communities in order to detect effects of Sure Start. If, after taking into account any fundamental differences between the 150 Sure Start communities and the 50 Sure Start-to-be control communities, differences exist in the functioning of children, families, and communities in the two sets of communities, these would be indications of the efficacy of Sure Start.

The second control strategy involves relying upon economically disadvantaged families being studied as part of the Millennium Cohort Study. A wide range of measurements of children and families to be used in the evaluation of Sure Start will be identical to those used in the MCS to ensure that comparisons can be made between economically disadvantaged children and families living in Sure Start communities (and studied by NESS) and comparable children and families not living in Sure Start communities (and studied by MCS).

One issue concerning the validity of comparisons between disadvantaged children and families studied as part of the MCS and the NESS is the impact of the concentration of local poverty. More specifically, is it possible that growing up poor in a community with a high concentration of poor families has different developmental consequences than growing up poor in a community with a lower concentration of poor families? For example, if the need for services is positively related to individual levels of deprivation, then a higher concentration of deprivation at the local level creates a greater pressure on local services and, in turn, might impact on the services which local individuals receive. Another example might be where peer pressure influences vary in relation to how many of your neighbours are similarly disadvantaged. Thus it is important that any such confounding effects are accounted for when comparing outcomes for NESS children and families to those of the MCS control children and families because the concentration of poverty is likely to be higher in Sure Start areas. One approach would be to estimate directly the effect of local poverty concentration for NESS and MCS children using, for

example, the index of multiple deprivation, and if such an effect exists, adjust for it when comparing outcomes of cases and controls.
Using multilevel models it will be possible to estimate separately:

- The effect of individual-level poverty
- The effect of community-level poverty
- The interaction between the two

The third estimate is known as a cross-level interaction and should offer insights into how the effect of individual characteristics is modified by the composition of the local community. For example do the results of any analysis differ for the four subgroups:

- Poor children in poor areas
- Not poor children in poor areas
- Poor children in not poor areas
- Not poor children in not poor areas

Of course, in addition to poverty, it will be possible to explore other individual and community characteristics, such as unemployment, and how these interact i.e. community effects moderating the effects of individual level variables upon individual outcomes. This kind of analysis can be done most immediately with variables that form part of the Index of Multiple Deprivation (IMD) as these are readily available at the level of community as specified by a digitised boundary or electoral ward. When 2001 census data becomes available (sometime in 2003), other variables can also be used in such analyses involving individual and community level effects. Such additional variables would include ethnic composition, household structure, occupational structure, overcrowded living conditions, and other aspects of demographic profiles. In addition to these variables, census data would allow variables derived from flux in population structure over the last decade (1991 to 2001) to also be included as a community level variable.

The MCS is a flexible resource from which to sample the controls for the Impact module of NESS. According to the research question under study, it will be possible to sample communities, and so the children nested within them; to sample single groups of children according to their or their family's characteristics; and, in some cases, where appropriate to use the entire MCS sample as a comparison group.

III. Selecting Ages of Measurement

The first data collection point should be in infancy to derive developmentally relevant information during a time of rapid developmental change. This developmental stage provides the foundation for later patterns of development and starts children on a particular developmental trajectory. The first months of life are difficult for collecting detailed interviews as many mothers will be making major adjustments in their lives and would find interviews unduly onerous. However, an interview when the child reaches 3 months of age is a good compromise in that it allows for a period of postnatal adjustment and is close enough to birth to maximise recall and birth experiences. Hence, 3 months old sets a bottom limit to ages for infancy contact. After this age a wide range of ages become possible for the infancy contact. As the child gets older, the recall of pregnancy and birth experiences by the mother becomes possibly a little less accurate. However, the mother is also gaining more experience of parenting and her child's characteristics and she can therefore report on these areas more fully.

The MCS chose 9 months as the age for their infancy contact. This age carries the advantage that the parents have undergone very substantial adaptation to the parenting role and can report on this. Also, parents will have a substantial body of experience of the child's characteristics. There is likely to be a somewhat less effective recall of antenatal and birth events than would be the case with an early infancy visit. The MCS decision was taken for largely fieldwork-pragmatic reasons. However, once this decision is taken, any study wishing to produce comparable data to MCS must adopt the same age points for data collection. Developmental change is too rapid in this period to countenance any variation in measurement ages across studies. The potential problem of poor recall of antenatal, birth and early postnatal events will be ameliorated where we can obtain individual data from health records, and this is currently being investigated.

Given that the infancy contact is taking place at nine months, which ages should be chosen for subsequent contacts? The decisions taken will be partly based on pragmatic issues such as fieldwork scheduling and partly theoretical developmental issues. The second contact should occur after a reasonable amount of time has elapsed for both developmental progress and family change to occur, but not so much time should elapse that the families forget about the study.

Around the age of two children undergo fundamental developmental change. In Piagetian theory this change is referred to as children moving from the sensori-motor to the pre-operational stage. This transition is accompanied by a dramatic change in the role of language in the child's development. A consequence of such a rapid developmental shift is the relative unreliability (instability) of measurement at this time. This would indicate advantages to having a second contact point later than 24 months of age. This developmental perturbation will have largely settled down by 30 months so that 30 – 36 months would be a good point for a second contact from a developmental perspective. However, from a fieldwork-scheduling perspective the 36-month contact would ensure that there was no overlap between the fieldwork for the 1st and 2nd contact points in the longitudinal study. This avoids problems whereby the amount of fieldwork increases temporarily requiring additional fieldworkers just for short periods. The advantages have led to 36 months being chosen as the second contact point in a longitudinal study.

For the third contact point, at least a year should be left after the second contact point in order for significant change for children and families to occur. This means in practice that 48 months onwards is an appropriate time for follow-up. If the gap between the 2nd and 3rd contacts is 2 years in duration, then this ensures no overlap for the fieldwork for the 2nd and 3rd contacts, which is a distinct advantage. Moreover, in England, almost no children will have started primary school at 48 months of age but 96% will be receiving some kind of pre-school centre experience. By 60 months of age most children will have started primary school. At 54 months around 40% will be in primary school. Hence there are advantages to choosing either 48 months or 60 months in that almost all children will be receiving equivalent experiences at the time of the contact. Taking all these considerations together the sequence of 9, 36 and 60 months of age for the timing of the first 3 contacts has some advantages. The advantages and disadvantages of these contact points are summarised below.

- There is no overlap in fieldwork between contact points
- There are roughly equal periods between measurement occasions
- Children are generally receiving equivalent patterns of home, preschool and school experiences at the same contact points
- There is no developmental reason for not choosing this combination or for choosing another set of ages in preference to this set
- A possible disadvantage of these ages is that they are outside the 0 – 4 age range usually ascribed as the target age range for Sure Start. However, in evaluating the effectiveness of Sure Start the demonstration of effects that are present beyond 4 years of age is advantageous, especially as the start of schooling might be regarded as a better demonstration of effectiveness of Sure Start than showing effects up to the age of 4. Also, children at ages 3 and 5 will have experienced Sure Start for a greater proportion of their lives than the same children at ages 2 and 4, hence the opportunity for Sure Start to have demonstrable effects is greater

The MCS team has indicated that they are agreeable to the 2nd and 3rd contact points of their study being at 3 and 5 years of age. The ESRC have also indicated that they wish there to be comparability of data between the National Evaluation of Sure Start and the Millennium Cohort Study.

If the ages of contact for the 2nd and 3rd contacts are the same for the cross-sectional study and the longitudinal study, then this offers some benefits in terms of efficiency of staff training and also in terms of allowing data to be compared across cross-sectional and longitudinal samples. For example in the cross-sectional study the points of contact could be 9, and 36 months, whereas the longitudinal study could have contact points at 9, 36 and 60 months. If this strategy were to be followed, then all contact points for the cross sectional study would be within the ‘reference frame‘ for Sure Start, and the longitudinal study could maintain comparability with the MCS and evaluate longer term effects for Sure Start at the start of formal schooling. It would also be the case that the 36-month data could be compared between cross-sectional and longitudinal samples should that be desired. For these reasons the ages for contact in the cross-sectional study have been chosen as 9, and 36 months and for the longitudinal study 9, 36 and 60 months of age.

IV. Methodology of Cross-sectional Study (150 Sure Start, 50 Sure Start-to-be [Control] Communities)

The second stage of the impact evaluation is a cross-sectional investigation of families in 150 Sure Start and 50 Control communities. Also it is the first stage of a longitudinal investigation for 100 Sure Start communities and families with infants in those communities. The presentation of the second stage of our impact evaluation is in three parts. First we specify how we will select control communities; next we detail how we will identify study children and families within Sure Start and control (Sure Start-to-be) communities; and in the third subsection we indicate what we will study with respect to individual children and families.

A. Identifying Control Communities

The Sure Start initiative is to expand by an additional 250 communities to a total of 510. The second 250 of these communities will be identified in 2001/2002. We will randomly select 50 of the 250 Sure Start-to-be communities (Waves 5 and 6) to act as control communities in stage 2 of the impact evaluation. For these control communities digitised boundaries will be created. With such boundaries, we will access the same administrative data sources that are central to the local context analysis to generate profiles of these control communities. Thus we will be able to characterise control communities in terms of geographic, demographic, ethnic make up and indicators of deprivation. Further details of this work are provided in the methodology report dealing with the local context analysis.

We now proceed to delineate our approach to the cross-sectional study of families with infants, and 3 year olds, in 150 Sure Start communities and 50 control (Sure Start-to-be) communities carried out in 2003 and 2004.

B. Sampling of Children and Families

In order to identify families with nine-month-olds, and three-year olds in Sure Start and Sure Start-to-be communities, we have made arrangements with the Department of Work & Pensions, Analytical Services Division Information Centre. Working from post-codes supplied by NESS, they will identify families in the targeted age range that receive child benefit. Nationally around 95 per cent of parents claim child benefit. In Sure Start and Sure Start-to-be communities the take up is likely to be effectively 100 per cent because the sums involved (especially the £15 a week for the first child) represent an important part of family budgets. A sample of parents based upon our specifications (i.e., child age and number of families) will receive a letter notifying them about the research and giving them the opportunity to opt out of the study. Other researchers have found that very few parents choose not to take part. Importantly, this is the very same procedure that the Millennium Cohort Study is using. Families participating in the MCS or any other study will not be contacted, the effect of such exclusions upon representativeness of the NESS sample should not be significant as other studies including MCS are using random sampling procedures. However as NESS will have available a range of demographic variables on the total population of families with a specific age child receiving Child

Benefit, the study can check the representativeness of the sample against these data at regular intervals in order to validate our sampling procedures.

This sampling strategy has been used successfully in the past by other researchers. One recent example was the 1999 Office for National Statistics study of mental health of children and adolescents in Great Britain. In that work, which included a wider range of age groups and a more varied social structure than Sure Start communities, only 6% of families chose not to be interviewed.

One challenge to the representativeness of the sample using the proposed procedures is the existence, within selected communities, of families with appropriate age children who are not registered for the receipt of Child Benefit. For the indigenous population the level of non-registration for Child Benefit should be extremely small as indicated above. However asylum seekers are not registered for the receipt of Child Benefit. Many of them receive support from the Home Office. Where such families exist within a community they would not enter our sampling framework and hence be omitted from the research. The Home Office keeps a central register of all asylum seeker families. The possibility of using this register as a means of sampling asylum seeker families, where they live within communities under study, is currently being investigated.

Another potential problem with this sampling procedure concerns families who move in the time period between registering for child benefit and being recognised as a potential participant in the study. This problem is unlikely to be of significance for families paid via a benefit book, as they are likely to register their new addresses in order to keep receiving their child benefit payments. However, registering of new addresses may well be slower for families having child benefit paid directly by credit transfer into a bank, building society or post office account, as these families would not have child benefit payments interrupted by not registering a new address. An analysis of child benefit payments to families in the first two rounds of Sure Start communities, using the child benefit database indicates that 10-20% of families in these areas are receiving child benefit by credit transfer. This only becomes a problem if these particular families change address in the time between registering for child benefit and being recruited to the study. We are currently investigating possible solutions to this issue by investigating the extent of such changes of address with some Sure Start communities who have good alternative data on families with a child less than four living in their area.

Period of sampling within communities

Families within Sure Start and Sure Start–to–be communities will be sampled over a 12-month period to achieve the target number of 9 month olds, and 3 year olds required to evaluate the impact of the initiative. A 12-month sampling period is needed, as a sample drawn from a shorter period may not yield enough 9 month olds for the study, in particular the longitudinal research. In addition, there are some slight seasonal variations in families' and children's characteristics and a 12-month sampling period overcomes any potential bias that would be introduced by seasonal variation if a shorter sampling period were used. For these reasons, the staged sampling of Sure Start communities over 12 months is planned.

The concern that, within communities, those children sampled later in the year will have had longer exposure to Sure Start services and are therefore not a comparable group with those sampled earlier in the year is acknowledged and the potential effect of such differences in exposure can be modeled in the analysis. Sampling within Sure Start and Sure Start-to-be communities over the same12 month period will mean that any seasonal or time of measurement factors will affect data for Sure Start and Sure Start-to-be communities equivalently and not be a source of bias for comparisons.

C. Detecting Effects on Individual Children and Families

Central to this part of our impact evaluation is a study of nine-month-olds, and 3- year-olds and their families in Sure Start and control (Sure Start-to-be) communities. We will **randomly** select and recruit the families of nine-month-olds, and 3- year olds, into our study, being sure to fully inform parents about the nature and purpose of the research and to secure informed consent. One hundred (100) children and families in each of 150 Sure Start communities will be studied in the cross-sectional phase of investigation; 80 with a nine-month-old, and 20 with a 3- year old. In addition 25 families with nine-month-olds, and 25 families with 3 year olds will be randomly selected and recruited from 50 control (Sure Start-to-be) communities.

For 100 of the 150 communities in the cross-sectional study, the 80 families and children seen at 9 months will be followed longitudinally at 3 and 5 years of age. This will be the longitudinal study within NESS.

The data obtained on these children and families will come from a variety of sources, though most notably from home visits carried out by a specially trained data collection team. Some data, as indicated below, will also come from medical records and from local Sure Start programmes.

D. Data Collection and Measurements

Given the goals of Sure Start, we have planned an extensive data collection to assess children's health and well being, that of their parents, and the functioning of their families, as well as their utilisation of services within their communities and their perceptions of their communities. Table 3 delineates the measures we plan to use or measurement domains we plan to study by means of a single $1^1/_2$-2 hour visit to each home. This visit will involve an interview with the principal caregiver of the target child, in most cases the mother. This measurement plan has been and will continue to be subject to revision pending (a) consultation with the Millennium Cohort Study investigators and (b) the results of piloting. All instruments used in both the cross-sectional and longitudinal studies were piloted in the autumn of 2001. Instruments were revised accordingly and, if need be, replaced. Because of the time demands that our data collection will place on families, we plan to offer them, in exchange for their time and effort, £10.00 coupons for use in local shops. All measurements proposed have been approved by the Research Ethics Committee at Birkbeck, University of London and the Multi-Region Research Ethics Committee (MREC) to ensure that our procedures follow standard practices for protecting the rights of research subjects and informing them of any potential risks associated with data collection, as well as their freedom to refuse participation in any aspects of the proposed research.

Following receipt of the MCS interview and questionnaire instruments for the 9-month contact, the NESS team drafted a composite set of instruments for the 9-month contact with children and families for the cross-sectional and longitudinal studies within the impact evaluation. The interview was then revised to be suitable for use with parents of 3 – 5 year olds. The range of research topics are presented in section D.1 in terms of those items taken from the MCS and those items that we are providing additional to the MCS set. Following the description of the interviews there is a description of the other content for the 3, and 5-year contacts.

Table 3

Individual/Family Data Collections

A. Family Demographics

1. Family and household composition
2. Income and source
3. Employment: mother, father, other
4. Education: mother, father, other
5. Marital status
6. Family debt burden
7. Ethnicity and religion

B. Family Dynamics (nonparenting)

1. Intimate relationship quality
2. Domestic violence
3. Parenting stress

C. Parental Health*

1. Mental health
2. Physical health
3. Health behaviour

D. Parenting/Home Environment

1. Home environment: HOME and children learning activities
2. Parent-child relationship
3. Parental discipline
4. Household order and organisation
5. Father contact and involvement
6. Parent literacy

E. Child Care

1. History: when, how much, who, where

F. Community

1. Observed local environment quality (checklist & observation, Barnes, 1997)
2. Housing adversity
3. "Belonging" to neighbourhood
4. Social support density of acquaintanceship, local social ties
5. Perceptions of local danger/disorder

G. Services Utilized by Children and Parents*

1. Ante-natal services used
2. Visits to GP over recent period and who for
3. Use of A&E and specialist services, e.g., mental health, including inpatient

4. Contacts with relevant agencies, including voluntary sector
5. Views about volume, quality, and appropriateness of services used
6. Whether service providers dealt with problems or referred them
7. Speed of response to a problem, and follow-up continuing care

H. Economic Context
1. Current/recent training/education activities of parents
2. Benefits family is receiving (e.g., free school meals)

I. Child Development **
1. Birth
    a. Birthweight
    b. Gestational age
    c. Maternal pregnancy and delivery history, including smoking, screening, and results, attendance at antenatal classes, labour/delivery complications
    d. Perinatal and infant health
    e. Type of feeding and duration of breast-feeding
2. Three years
    a. Social/behavioral development: Bayley Scale Ratings (observation)
    b. Social/behavioral development: (parent report)
    c. Language: subscales from British Ability Scales
    d. Non-verbal skills: subscales from British Ability Scales
    e. Cognitive Development: British Ability Scales
    f. Health: feeding, height, weight, mobility, history of injury and illness; developmental delay; immunisations; congenital conditions
3. Four and five years
    a. Social/behavioral development: Bayley Scale Ratings (observation)
    b. Social/behavioral development: (parent report)
    c. Language: subscales from British Ability Scales
    d. Non-verbal skills: subscales from British Ability Scales
    e. Cognitive Development: British Ability Scales
    f. Emergent Literacy: pre-literacy scales from EPPE
    g. Health: height; weight; arm circumference; immunisations; diet and nutritional status; special needs; hospital admissions

* Additional data on service use may be available from programme monitoring data and these data can be used in between Sure Start comparisons. However they would not be available for Sure Start versus non-Sure Start comparisons.

** A more extensive list of child and parent health measures may be obtained from health record systems (e.g., Hospital Episode Statistics, Incapacity Benefit, NHS Trusts, Disability Living Allowance), where parents give permission so that we can identify individual parents and children in the aggregate record archives. The appendix to the Local-Context-Analysis methodology report delineates these measurements, some of which are Sure Start targets.

All interviews, questionnaires and related measurements will be translated for use with non-English-speaking families. Diverse staff will be hired to increase the comfort that minority families in particular experience during home visit and phone-call data collections. As a last resort, translators will be hired to assist data collectors when gathering information from families with whom project staff do not share a common language. Where parents have problems with literacy, or have not read any of the project leaflets provided for them, the information concerning their options to opt-out of the study and the description of the study and the obtaining of their informed consent will be fully explained to them verbally so that these parents are able to participate on an equivalent basis with all other potential participants. Finally, when it comes to testing children, children who are not fluent--relative to their developmental status--in English will be administered only nonverbal tasks. All these data collection strategies have been used successfully in the EPPE project.

1. The Interview to be used with Parents of Nine-month Olds

The interview will take place with the principal caregiver, who will usually be the mother. In order to achieve a high level of comparability of data with that collected within the Millennium Cohort Study (MCS), we have chosen to use the same items as exist within MCS wherever we judge that sensible. This point also applies to 3, 4 and 5-year measures. Hence in this interview some topics will overlap with topics in the MCS. For these topics we have used the same questions and coding as the MCS in order to enable easier comparison of data between the two studies.

Direct overlap with MCS

The topics where there is direct overlap with MCS are as follows:

- Household details including marital status, ethnicity, father-relationship to baby and mother
- Lone parenting
- Pregnancy details including planned/or not, prenatal classes, health problems, labour and delivery, hospital location, delivery complications
- Baby's health including breast feeding, red book consultation, hospital admissions, immunisations
- Baby's development including motor, communicative, sleeping/crying
- Childcare
- Parent's own family history including experience of childcare
- Friends and support
- Mother's health including mental health, physical health, treatment, smoking and alcohol use
- Mother's employment
- Partner's employment
- Mother's current/recent job (SES classification)
- Partner's current/recent job (SES classification)
- Family income and benefits
- Mother's and father's education (highest attained)

- Mother's literacy and numeracy
- Housing and facilities
- Neighbourhood characteristics
- Service utilization
- Interests/religion/associations/groups

(Self-completion in MCS: to be administered in interview format in NESS)

- Baby temperament
- Parenting beliefs
- Mother-baby relationship
- Mother's malaise (mental health index)
- Marital relationship
- Attitudes to neighbours
- General attitudes
- Self-esteem (psychological state index)

Additional measures to those used in MCS
In addition to the above, NESS will be including additional topics in the parental interview that are not in the MCS:

- Household order and organization: chaos, hubbub and order scale
- Childcare history (MCS has some data but NESS will collect more extensive data)
- Service use (MCS has some data but NESS will collect more extensive data)
- Parent-child relationship scale
- Discipline: parent limit setting
- Parent-child conflict tactics scale
- Substance use and abuse
- Stressful events
- Domestic violence and marital harmony: sections on negotiation, psychological aggression and physical aggression from conflict tactics scale

In addition to the above topics dealt with in the parental interview, fieldworkers will complete the following observation scales around the home visit.

Brief Neighbourhood Observation (Barnes, 1997)
HOME Observation (Emotional Responsivity, Cognitive Stimulation) (Caldwell & Bradley, as used in Early Head Start Study)

2. 3, and 5 Year Old Contacts
Parental interviews/home visits
In addition to topics covered at 9 months, the following additional interview/observation instruments will be used.

Parental Aspirations for the Child (Anning, 2000)

Activities with Children (Home Learning Environment (HLE) from EPPE)
Child Behaviour Rating During Assessment (Bayley, 1969)
Environment During Assessment (Bayley, 1969)

3. Childcare and Preschool Environments
In the 100 communities where we are studying families and children longitudinally, we will also be investigating the major forms of childcare and pre-school provision available to the children in the community. For children in our impact study, where they attend a childcare or pre-school centre ten hours a week or more, we collect data on their main childcare or pre-school centre. These data collections will be via interviews with facility staff and observations of the children's environments using the Early Childhood Environment Rating Scales – Revised (ECERS-R) (Harms, Clifford & Cryer, 1998) and the Early Childhood Environment Rating Scales – Extension (ECERS- E) (Sylva et al. 2000) and the Adult-Child Interaction Scale (Arnett, 1989). These instruments have been used on the EPPE and NICHD projects and have proved useful in elucidating how childcare and pre-school facilities may contribute to children's learning and development.

4. Child Assessment at 3 and 4 Years of Age
The British Ability Scales (BAS) will be used as measures of cognitive and language development. These scales are standardized from 36 months of age upwards, and provide measures of non-verbal and verbal (language) abilities from four subscales, block building, picture similarities, picture naming and verbal comprehension at 36 months. Other subscales become available for older children.

At 3 years of age the Strengths & Difficulties questionnaire supplemented by additional items developed within the EPPE study will be used to assess positive and negative aspects of social/behavioural development. This parent-report instrument provides measures of several aspects of children's social development.

5. Child Assessment at 5 Years of Age
At 5 years of age British Ability Scales (BAS) subscales and literacy measures from the EPPE study will be used as measures of cognitive, language and school readiness development. The subscales will be:

- Verbal comprehension
- Picture naming
- Picture similarities
- Pattern recognition
- Early number skills

In addition to the BAS subscales the following measures of emergent literacy, as developed in EPPE, will be used:

- Alliteration
- Rhyming
- Letter recognition

This combination of measures can be combined in several ways to produce measures of:

- Overall cognitive development
- Verbal cognitive (language) development
- Non-verbal cognitive development
- Phonological awareness
- Emergent literacy/ book behaviour
- Mathematical/numeracy thinking

For social/behavioural development, NESS will use the Strengths & Difficulties Questionnaire (SDQ) supplemented by additional items from the EPPE study (see above).

At 3, and 5 years of age for the longitudinal study we are considering additional measures of emotional development. In collaboration with the ECLS we will be field-testing a specially developed Q-sort for the measurement of child attachment security. This Q-sort has been specifically developed for large-scale longitudinal studies as exist in ECLS and in NESS. If judged appropriate after field-testing, the Q-sorts may be included at 3, and 5 year contacts, and possibly adapted for measuring attachment security to non-parental caregivers as well as parents.

6. The Learning and Educational Aspects of Children's Environment and Development

The extensive investigation of the home background of children through the measures discussed above in relation to home visits at 9 months, 3, and 5 years of age will enable us to examine the influence of the home environment on children's development. Additionally data on the neighbourhood/community, the nature of Sure Start programmes and the childcare/pre-school environments will elucidate the out-of-home environment as it is relevant to children's learning and development. The child assessment measures at each stage of the impact study will cover cognitive/language and social/behavioural aspects of development that are central to the children's developing school readiness. Hence the impact study will clearly be producing information on the role of home and out-of-home environments (including aspects related to Sure Start programmes) in the development of the children's educational preparedness. It would be useful in the long-term to investigate the relationship of these measures to educational measures taken at the start and end of the foundation stage and the end of Key Stage 1.

V. Third Stage: Longitudinal Study of 100 Communities, their Children and Families

The third stage of our impact evaluation is organized around the longitudinal study of children, families and communities using 100 Sure Start communities. Eighty families will be longitudinally studied in each of the 100 Sure Start communities. The families to be studied were those studied on a cross-sectional basis in measurement year 2. Families in the 100 communities chosen for longitudinal study will be re-contacted when their child is 3 years of age (2005 for TB/Wave 1, 2006 for Waves 3 & 4) for a long-term longitudinal investigation of child and family development. If families have moved out of the communities in question, they will be followed wherever they have relocated, provided it is within the UK. This will enable us to ask and answer dose-response questions about the impact of Sure Start, contrasting children and families that were exposed to Sure Start (i.e. lived in Sure Start communities) for longer and shorter periods of time. In the initial contact with families and in all subsequent contacts, we will secure the name, address, and phone number of at least two contact people who would likely know the whereabouts of a study family were they to relocate. Additionally, on an annual basis, we will update family addresses and phone numbers through mailings, as well as information on the employment and education of parents. These strategies have been used successfully in the past to track mobile families.

When the children are 3 and 5 years old, all the age-appropriate, individual and family measurements listed in sections C.1.2 and C.1.3 above will be collected and aggregate community level data will be collected yet again. With these data in hand, it will be possible to address three questions (i.e., **Are there differences between Sure Start and Control communities? Does the efficacy of Sure Start vary as a function of programme strategy or characteristics? Are apparent effects of Sure Start actually a function of other government initiatives?**) using multivariate, longitudinal data analytic techniques. It will be possible to determine whether children from Sure Start communities show more beneficial development (physically or psychologically), than children from control communities; whether income, parenting, housing or neighbourhood characteristics operate differentially for families in these two types of communities; and whether any detected effects should be attributed to other government programme initiatives rather than to Sure Start or to differences in mainstream services (as measured within The Local Context Analysis module). The multilevel analysis will also identify those Sure Start communities that produce the most change. Similar strategies underlie answering these questions whether the dependent variables relate to children, families or communities. These analyses can address dose-response issues with regard to exposure to Sure Start or other initiatives or programmes. In addition to overall Sure Start exposure, the dose-response analysis can be extended to particular types of services that might be used by families within communities.

Are There Differences Between Sure Start and Control Communities?
The 50 Sure Start-to-be communities provide one type of comparison. Comparisons between Sure Start and Sure Start-to-be communities provide a basis for evaluating the effect of Sure Start. This will occur initially when the children are infants. However as the children age, the data on infants and 3 year olds and their families in the longitudinal Sure Start communities can be compared with the data on same age children and their families who were seen in the Sure Start-to-be communities in the cross-sectional study. We should note that these latter kinds of comparisons have definite limitations and for this reason they do not form the principal basis for our examination of the effects of Sure Start. A big limitation of these types of comparisons for 3-year-old contacts is that they involve comparing data collected on a first occasion (in the Sure-Start-to-be sample) and on a second occasion in the longitudinal Sure Start sample. As repeated assessment can exert its own effect on measurements, these comparisons will have to be interpreted

cautiously. Hence Sure Start-to-be communities do not provide the principal control group for the longitudinal study.

Nevertheless, the same types of multivariate analyses described previously can be undertaken. The hierarchical structure of the data means that multilevel modeling is appropriate. This technique controls for the inclusion of variables measured at different levels of the hierarchy i.e. both child-specific and community-specific measurements. In particular, whether each community is a Sure Start community or not can be thought of as a community-specific variable. We can contrast Sure Start and non-Sure Start communities, including covariates to control for pre-existing differences between communities. We intend to include in these analyses measures of the quality of mainstream services in health, social services and education as developed within the Local Context Analysis module. We will also build into these analyses, as appropriate, contrasts between regions of the country, ethnic variation, and other factors that may moderate the differences between communities.

As already noted above, the children and families included in the MCS study who are economically disadvantaged and do not reside in Sure Start communities will serve as another type of control. This will be the principal control for Sure Start versus non-Sure Start comparisons for the longitudinal study. Where measurements and children's ages are equivalent, the data from these MCS children and families will be compared with data from Sure Start children and families, using similar analytic strategies to those previously described, with one important addition. As noted previously in the discussion of control and comparison groups, the effect of growing up in a poor family under different concentrations of poverty in the community will be explored using the MCS data and any comparisons can be adjusted for such effects.

Once these adjustments are made, statistical models used in analyzing the longitudinal data will take the repeated-nature of measurements into account. In our multilevel model we can think of this as another level of the hierarchy, with measurement occasions (at level 1) being nested within individuals (at level 2) who live in communities (level 3). This will enable the correct analysis of factors which are constant across measurement occasions at the level of both child (e.g., sex) and community (e.g., region of the country) as well as those which vary from one measurement occasion to another at the level of child (e.g., age) and community (e.g., employment rate). The longitudinal analyses will enable the consideration of the mathematical function relating outcome to years of Sure Start; e.g. whether increasing years of Sure Start have increasing, decreasing or plateau effects.

Does the Efficacy of Sure Start Vary as a Function of Programme Strategy or Characteristics?

The longitudinal sample of 100 communities will include the major forms of Sure Start strategy. Hence multivariate longitudinal data analytic techniques will enable discrimination of differences related to programme strategy or characteristics. Community strategy and characteristics will apply at the community level of analysis with children and families nested within communities. The longitudinal nature of the data will enable different delays of effects for different strategies to be elucidated.

Are Apparent Effects of Sure Start Actually a Function of Other Government Initiatives?

Once again to ensure that effects from between- and within-group analyses attributed to Sure Start are indeed a function of Sure Start, follow up analyses will be carried out to determine whether the presence of other government initiatives in communities are likely to be responsible for detected Sure Start effects. As before, analyses will also determine whether Sure Start interacts with other programmes in their impact on communities.

Sub-group analysis for different sub-groups is also possible, e.g. major ethnic groups, lone parents, teenage parents, refugee/asylum seekers (if sampling proves possible). Where such sub-groups are larger than 400 then there is considerable power to detect small effects. With sub-groups smaller than 400 the power will be critically dependent of the size of effects. The longitudinal sample size has been chosen so that it is anticipated that sub-groups of major relevance in social policy formation will be present in groups greater than 400.

The collection of longitudinal data within the first cycle of funding at the individual level on children and families will enable us to expand upon the two questions mentioned above that are central to both the cross-sectional and longitudinal studies to be included in our impact evaluation. More specifically, we will be in position at the end of the first funding cycle to answer a third important question. **Does exposure to Sure Start (or to particular types of Sure Start programmes) deflect and modify developmental trajectories in ways that protect or buffer children and families from otherwise anticipated risks?** Extensive scientific evidence shows that many characteristics of socially excluded children living in deprived circumstances (e.g. low birthweight) and families (e.g. limited parent education) predict a host of problematic outcomes for children and families alike. We will test the hypothesis that linkages between early indicators of risk and later functioning will be attenuated in the case of Sure Start children and families relative to Millennium Cohort Study control children and families. Thus, we will determine whether factors like low birthweight and limited maternal education, to cite but two examples, predict poor cognitive functioning of children and harsh parenting at the time children are 3 years of age equally in the Sure Start and Control samples of families. To the extent that Sure Start modifies the developmental trajectories of children and families as it aims to do, then we can expect that predictive associations of the kind under consideration will be attenuated in the case of Sure Start communities**.** In addition to addressing between-group comparisons between Sure Start and Control samples using multiple regression techniques or structural equation modelling procedures, the same statistical techniques will be used to address within-group questions contrasting different types of Sure Start programmes (e.g., those that emphasised health vs. those that emphasised education).

VI. Fieldwork Plan

A. Sampling
A monthly cycle of sampling will be enacted. The first of these repetitive monthly cycles is described here.

During the last week of October 2002 the DWP will receive the postcodes for the areas to be sampled.

During the first week of November 2002 letters about the study will be sent to families whose children will be aged nine months, 3 and 4 years of age at the time of the interview. Families will be given the opportunity to opt-out of participating in the study.

During the 3rd and 4th week of November 2002 the DWP will check the sample for sensitive cases and will also check whether families are involved in another study, and have not opted out of participation.

During the 1st week of December 2002 the names and addresses of the families taking part will be released. Families will be selected randomly from this list.

During the 2nd week of December 2002 fieldwork will be allocated to the regional fieldworkers.

January 2003- Fieldwork commences.

During January to December 2003, 75 Sure Start communities from rounds 1 and 2 and 50 Sure Start-to-be communities from rounds 5-6 will be visited. During January to December 2004, 75 Sure Start communities from rounds 3 and 4 will be visited. This means that we have a random sample of nine-month-olds, and 3 year olds and their families for 150 Sure Start communities and 50 Sure Start-to-be communities where sampling within any community occurs over a one-year period.

B. Fieldwork Management
The fieldwork for visits to families with a 9 month old has been subcontracted to the Survey Division of the Office for National Statistics (ONS). ONS is a partner in the National Evaluation of Sure Start and is also contributing to other modules within the evaluation. They have the most experienced survey team in the UK and have conducted surveys on a large number of topics, including very sensitive topics, with all sections of society, including hard-to-reach, deprived and socially excluded populations. There is close liaison between the ONS and NESS teams so that the work of the two teams is closely integrated.

The project coordinator will coordinate the work of regional supervisors and their teams under the supervision of the Impact Study director, Professor Melhuish. Each regional supervisor will be responsible for ensuring that 3-5 fieldworkers are adequately trained and checked for reliability throughout the data collection period.

1. Training
Training for all fieldworkers/supervisors will take place in September/October. Everyone using the British Ability Scale (BAS) and the pre-literacy scales will be thoroughly trained and assessed for reliability, and colleagues from the Office for National Statistics (ONS) will conduct a training course on general interviewing procedures and using the Computer Assisted Personal Interview (CAPI) schedule. Training will initially take place in London and will be followed up regionally by fieldwork supervisors.

2. Reliability
To assess that fieldworkers are fully trained in every aspect of the interview, child assessment and home visit, it will be necessary to ensure that all data are collected and visits completed in a reliable and consistent manner. Therefore, a central fieldworker will take on the role of reliability assessor, thus acting as a 'gold standard' interviewer for all fieldworkers.

The central assessor will ensure that all fieldworkers are proficient in carrying out every aspect of the home visit before they embark on full assessments in the field. She will complete a home visit with all fieldworkers to ensure that collected data are consistent. Reliability checks will ensure that assessments are in line with each other and that any necessary amendments are made.

3. Quality Control
To ensure that all interviews/assessments are fully completed and valid, data sent back to the central office will be thoroughly checked for missing data and inconsistent responses. All fieldworkers will be responsible for the quality of their data; where missing/contradictory data are found, families will be contacted and asked to respond to the relevant questions. Efforts will be made to ensure the time lag is not too great between interviews and any follow-up questions and that problems with data collection are the exception rather than the rule.

4. Piloting
Piloting will take place from September to November and in light of this the instrumentation will be revised by December 2002. The pilot study will focus on the appropriateness of methods to be used in the study. Sure Start communities that are not part of the main sample will be randomly selected for participation in the piloting study. This ensures that a similar population will be sampled. The piloting will include 20 interviews per region and fieldwork supervisors will closely examine collected data. As previously mentioned, the child assessments will be further checked using a 'gold standard' interviewer. The duration of interviews/child assessments and initial problems encountered in the field can be monitored using the information from the pilot study.

5. Panel (Sample) Management and Maintenance

The study aims to maintain 80% or greater of the families initially approached within the study. In order to maximise the likelihood that a family will agree to take part in the evaluation research a number of techniques and approaches will be employed. In most cases, it is expected that an initial phone call would be made to the family to seek an interview. The fieldworker 'personalising' the call, i.e., using their own name, the family name and the name of the child helps to facilitate rapport. Being sensitive to any background clues that indicate it might not be a convenient time to talk, e.g., a crying child or the sound of visitors and instead agreeing another time to call back encourages future co-operation. Reminding the parent/ carer of the letter inviting them to take part in a national survey, stressing that they have been chosen by chance, that they remain anonymous and that all data is held in the strictest confidence will help to reassure potential participants.

The telephone can be a difficult medium for communication and this may be especially so for parents/carers for whom English is not their first language or for those with disabilities. Where telephone contact is not proving to be productive, fieldworkers may offer to call in person to talk through the aims of the research and what would be involved. In addition, fieldworkers have the flexibility to organise interview times and locations to suit families' daily routines. If an evening or weekend appointment is requested, this can be accommodated. If parents/carers agree to participate but would rather the interview was not held in the home, fieldworkers have the authority to seek a suitable alternative venue.

Most parents enjoy the interview as many of the topics are inherently interesting to them. This acts as an incentive to continue in the study. An additional incentive is the £10 shopping voucher that all families will receive upon completion of each home visit.

To reduce attrition from the longitudinal sample, the contact details of two individuals known to the family will be requested, e.g., close friend or family member. If the family should move during the course of the research, these contacts will enable communication with the family to be re-established. Between visits, producing a short, cheerful newsletter about how the research is progressing reminds the family of their participation in a national survey which encourages them to feel 'part of' the research and more likely to want to continue their involvement. The study team will send birthday cards to each child and Christmas cards to each family in the longitudinal study and these also serve to give families a sense of involvement in the research project. There will also be annual follow-up telephone calls to families to keep in touch and to check for any changes in address or pertinent circumstances.