

# Assessing preschool children's emotional and behavioural adjustment

**A report back to the Department of Health on two projects involving the Development and Well-Being Assessment. Written by Robert Goodman on behalf of the youth*in*mind team**

## **Executive summary**

- 1) At a time when there is cross-party consensus on the importance of the preschool years, it is very unfortunate that there is so little national data on mental health in these years – in marked contrast to the excellent national data on the mental health of 5-16 year olds. The successful modification of the Development and Well-Being Assessment for children aged 24-59 months opens the way to filling this evidence gap. (“Infant psychiatry” has not yet developed to the point where inclusion of 0-23 month seems justified.) A nationally representative survey of the mental health of 4,800 children aged 24-59 months is recommended.
- 2) Online assessment can generate similar findings to face-to-face assessment, and do so at lower cost. However, this is accompanied by low parental participation rates, particularly in deprived areas, among those who rent their home, non-traditional families and British Asians. These features may undermine the credibility and acceptability of surveys that rely exclusively on online assessment. By contrast, offering respondents the choice of whether they wish to complete interviews face-to-face or online has the potential to increase response rates while also reducing the average cost of an achieved interview.
- 3) Physical health surveys are generally obliged by Ethics Committees to provide feedback to participants who are found to have treatable disorders that they are not aware of. The same is not true for mental health surveys. Preliminary evidence suggests that parents who have completed a survey about their child's mental health do welcome being offered immediate feedback on the information that they have just provided. The computerized feedback was generally perceived as helpful or neutral, but some respondents experienced it as impersonal or slightly upsetting. The findings warrant further investigation, but the evidence at present is not sufficiently strong to provide definitive guidance on when or whether feedback should be used.

*This report begins with a 11-page narrative summary that is designed to be understandable by everyone. It is followed by a longer technical report that is designed primarily for researchers, covering in much more detail the methods and samples used, the results obtained, and the inferences that can be drawn from these results.*

## **Importance of the early years – the policy background**

Irrespective of political party, successive UK governments have increasingly emphasised the importance of the preschool years. This has been evident, for example, in the promotion and funding of the Sure Start local programmes and their successor Children's Centres. Further evidence of policy makers' continuing interest in the early years is to be found in the recently published report of the Independent Review on Poverty and Life Chances, commissioned by the Prime Minister in June 2010 and led by Frank Field MP (Field, 2010). This review recommended that national and local government "should give greater prominence to the earliest years in life, from pregnancy to age five, adopting the term Foundation Years" – adding that one reason for this was "to establish the Foundation Years as of equal status and importance in the public mind to primary and secondary school years."<sup>1</sup>

The Field report highlighted the extent to which social inequalities in life chances are firmly in place by school entry, suggesting that efforts to provide all children with equal opportunities need to begin in the preschool years. The early years seem to be a time when – for good or ill – children are particularly sensitive to the quality of their cognitive, social and emotional environments. Stimulation, deprivation or inappropriate handling at this time can set individuals on different life courses that are subsequently relatively hard to shift. A society that wishes to maximize its well-being, its human capital and its national productivity would seem well advised to focus on its very youngest citizens.

A focus on the early years is often motivated by the desire to detect those young children whose development is going awry as soon as possible. Hopefully, this will occur early enough to make it easy to nudge these children back on course before they get caught up in vicious cycles of self-perpetuating failure, stigma, low self-esteem and disaffection. Detecting learning difficulties is one important part of any such program, and so too is detecting emotional and behavioural difficulties<sup>2</sup>. This is not only because emotional and behavioural difficulties are powerful barriers to learning and work, but also because emotional and behavioural difficulties make children, families and classrooms unhappy and can leave a lasting legacy in the form of children who grow up to be adults with higher rates of mental illness, criminality, unemployment and poor parenting.

This report is primarily devoted to one practical aspect of the greater attention that policy makers are now giving to the emotional and behavioural adjustment of young children – namely, the need for a valid and appropriately detailed measure of psychological adjustment in the under-5s. Such a measure is clearly essential if local and national surveys are to establish how common different sorts of emotional and behavioural difficulties are in the early years. The same measure is also needed for follow-up and repeat surveys to detect year-on-year trends, e.g. as a result of government interventions. If the measure can be applied not only for surveys but also for screening or clinical assessments, so much the better.

<sup>1</sup> Quoted from page 6 of the report.

<sup>2</sup> The terms "emotional and behavioural difficulties", "mental health problems", and "psychiatric disorder" are used from different professional perspectives to refer to very similar phenomena, and are used interchangeably in this report.

Previous British nationwide surveys of child and adolescent mental health – and why the early years were not included

The Department of Health (DH) has funded a rolling program of nationwide surveys of mental health, beginning with a survey of adult mental health in 1993 and subsequently involving new adult samples in 2000 and 2007. The program was expanded to include 5-15 year olds in 1999 and subsequently involved a new sample of 5-16 year olds in 2004. These surveys are internationally outstanding in terms of the size of the samples and the quality of the measures. The findings from these surveys have been of great use to those involved in planning, commissioning and providing mental health services.

It is a source of frustration to those with a particular interest in the under-5s that there is no comparably good nationwide information about young children. This omission was due to the lack of suitable measures of emotional and behavioural adjustment in under-5s at the time when the initial nationwide surveys were carried out. A feasibility study in 1997 tried to survey the mental health of 3 and 4 year olds using an adapted version of the Behaviour Checklist (Nicol et al., 1987), but the results of this feasibility study were discouraging, leading an expert group to recommend that under-5s be excluded from the subsequent nationwide survey because of the problems in finding an appropriately sensitive instrument (Meltzer et al, 2000).

### **A soluble problem – but an impractical solution**

At the time of the original feasibility study, the prevailing view was that preschool symptoms were best seen as free-standing problems that could not generally be grouped into diagnoses (i.e. recognised clusters of symptoms). Over recent years, this view has changed and international diagnostic systems have been extended to include preschool children. Helen Egger and her colleagues have developed a new preschool assessment of mental health, the PAPA (Preschool Age Psychiatric Assessment), as a downward modification of the well-established CAPA (Child and Adolescent Psychiatric Assessment) for older children. They have used this new tool to demonstrate that diagnoses that were previously thought to be restricted to school-age children can validly be applied to preschool children too – both in epidemiological and clinical studies (Egger et al, 2006). The PAPA is arguably the best preschool assessment ever developed. Unfortunately, it is not really suitable for nationwide surveys since it is a very long assessment that has to be administered by interviewers who are much more highly specialized and intensively trained than would be possible for the British organizations that are in a position to carry out national surveys. Furthermore, even if a nationwide survey using the PAPA were possible, it would carry a price tag that would almost certainly be unaffordable at present.

### **The Development and Well-Being Assessment – a practical solution?**

The successive surveys of child and adolescent mental health that were commissioned by DH and carried out by the Office for National Statistics (ONS) between 1997 and 2007 all used the Development and Well-Being Assessment (DAWBA) as their principal measure of mental health disorders. The DAWBA was developed by Robert Goodman from 1996 onwards as a mental health assessment of 5-17 year olds, partly to provide a

suitable epidemiological tool for the nationwide British surveys and partly to provide a useful assessment tool for clinics (Goodman et al, 2000a).

The DAWBA collects information from multiple informants, records open-ended descriptions of the problems, and generates diagnoses based on the expert judgment of an experienced clinician. Information is nearly always collected from the principal caregiver (typically the child's mother), supplemented where possible by information from teachers, and a self-report from the young person being assessed. Since self-reports are unreliable in younger children, assessments of under-11s rely just on parent and teacher reports. Whenever respondents report difficulties, they are prompted to describe those difficulties in detail in their own words – with these open-ended comments being recorded. Experienced clinical raters review the available information from all informants before rating which disorders, if any, the child has.

The nationwide British surveys based on the DAWBA are world leaders in scale (covering around 20,000 children and adolescents), breadth and clinical depth (Meltzer et al, 2000, 2003; Green et al, 2005). These surveys have served as a model for similar DAWBA-based surveys of 5-16 year olds in many other countries, including Bangladesh, Brazil, China, Denmark, India, Israel, Italy, Norway, Russia and Yemen.

*DAWBA innovation 1: development of the early years version*

Developed originally for the assessment of 5-17 year olds, the DAWBA has recently been modified to cover 2-4 year olds and adults. The preschool (or “early years”) DAWBA differs from the standard version for schoolchildren in several ways.

- a) Some sections of the 5-17 DAWBA have been dropped because the relevant disorders are so rare in 2-4 year olds that they are almost never encountered in specialist clinics, let alone in general population samples. The dropped sections covered: panic disorder and agoraphobia; obsessive-compulsive disorder; anorexia nervosa and bulimia nervosa.
- b) Other sections have been added to cover disorders that are particularly relevant to 2-4 year olds, namely: reactive attachment disorder (inhibited and disinhibited); feeding disorders; sleep disorders; and elimination disorders.
- c) Adjustments have also been needed for those sections that are relevant both in the early years and in later childhood since the characteristic symptoms may vary according to the maturity of the child. In the early years, for example, evidence of post-traumatic stress disorder (PTSD) is more likely to include repetitive play or drawing related to the trauma.
- d) At all stages of childhood and adolescence, disorders are defined not just by the presence of a recognized constellation of symptoms but also by substantial resultant distress or social impairment. In the early years, as in later childhood, it is relevant to consider social impairment in home life, friendships, learning and leisure. In the early years there is an important additional category of social

impairment that is routinely assessed by the DAWBA— lack of access to day care because of mental health problems.

While the early years DAWBA covers the mental health of 2-4 year olds, this does still leave the period from 0 to 23 months uncovered. Ongoing research on “infant psychiatry” may one day allow this important developmental period to be covered by high-quality survey instruments, but for the present this is an aspiration rather than something that can be delivered in the near future.

*DAWBA innovation 2: online administration for epidemiological surveys*

An online version of the DAWBA was developed by the youthinmind team, and has been widely used in over 20 languages by child and adolescent mental health clinics throughout the world. Though not all families have easy access to the Internet, an online assessment does provide a convenient, cheap and thorough means of assessing many children and teenagers. It is convenient for many families since they can fill in their answers at times that suit them, e.g. over the course of a number of evenings. They can report as much as they want, without the constraints of a brief appointment, and they can subsequently add to, or modify, their answers if they realize on reflection that their initial report was misleading or incomplete. An online assessment is cheaper because it dispenses with the need for an interviewer. It is also particularly thorough because computers, unlike clinicians, can be relied on to remember to ask a long list of potentially relevant questions. This preliminary data gathering by a computer facilitates rather than replaces the subsequent interview with a clinician. Once the initial data gathering has been completed online, the clinician can concentrate on engaging the family, resolving ambiguities and probing further where that seems indicated. An online assessment no more replaces the need for skilled human input than a hip X-ray replaces the need for an orthopaedic surgeon.

Though the online DAWBA has been used in clinics for many years, it is only recently that it has been used in epidemiological surveys. The best evaluated instance is in Norway. Bergen is Norway’s second largest city and its “Children in Bergen” surveys have monitored the mental health of a large cohort of children (N=10,000) on three occasions: when the children were aged 8-10 initially; when followed up 3 years later aged 11-13; and when followed up for a third time aged 14-16. On each occasion, the DAWBA has been the main measure of mental health. In the first wave, all DAWBAs were administered by interviewers, whereas in the second and third waves, DAWBAs were administered online. Response rates were about the same with and without interviewers (around 45%), while costs were reduced by about 75% by switching to online self-completion, even taking into account the full cost of inviting respondents and reminding them if appropriate (Heiervang and Goodman, 2010). However, the switch from interviewers to internet was not without problems. One important disadvantage of online administration was selective participation, with under-representation of less-well-educated mothers, single-parent families and immigrants. Another disadvantage of online administration was more missing data: interviewers were good at encouraging respondents to see the interview through to the end, whereas around half of the online respondents logged off before completing the entire interview.

## **DH commissions research into a practical solution**

Aware of the “evidence gap” left by the lack of nationally representative information on the mental health of preschool children, and aware too of the pressing need to “do more with less” in future surveys, DH commissioned youth*in*mind to evaluate the utility of the early years DAWBA. This evaluation was to be done via online surveys in such a way that the findings would simultaneously shed useful light not only on the early years DAWBA, but also on the practicalities of adopting online surveying for future national surveys of child and adolescent mental health.

While the DAWBA can collect information from parents, teachers and young people aged 11 or more, the commissioned studies only collected information from parents – partly because parents are the key informants for young children, and partly to keep down costs. The Norwegian experience with online DAWBA surveys has shown that applying the online DAWBA to teachers and teenagers involves the same benefits and challenges as applying that online assessment to parents. There is no particular reason to suppose that this would not apply equally in Britain.

## **Preparation for data collection**

Once the contracts were signed off in late 2009, youth*in*mind and its partner organizations intensified their preparations for field work. The online DAWBA was upgraded – not only to include the early years interviews, but also to offer parents who completed the DAWBA the opportunity to get immediate automated feedback based on the information they provided about their children. For the last decade, youth*in*mind has pioneered the provision of free, extensive and up-to-date information on children and teenagers’ stress, adjustment and well-being. Users of [www.youthinmind.info](http://www.youthinmind.info) are able to complete a very brief assessment (the Strengths and Difficulties Questionnaire (SDQ)), receive immediate feedback on that assessment and search the website’s directories to find relevant books, websites, helplines and local services.

Many of the readers of this report will be surprised to hear that the previous British nationwide surveys of child and adolescent mental health did not offer any feedback to respondents. It is, of course, heartening that so many families are willing to contribute a lot of their time to the surveys with no direct return, beyond the pleasure of having an interested listener and the satisfaction of contributing to the public good. Nevertheless, there is a case for believing that people who contribute to health surveys earn the right to individual feedback, particularly if a survey picks up a treatable health condition that they did not previously know about. Such feedback is the rule for surveys of physical health, and it somewhat anomalous that similar feedback is not routinely offered after surveys of mental health. When respondents express concerns, the interviewers in mental health surveys typically limit themselves to providing respondents with generic advice on who they could contact (e.g. their GP) without providing any individual feedback to respondents or their GP. For this study, the detailed online DAWBA assessment was melded with the youth*in*mind feedback.

Parents who completed the online DAWBA assessment were asked if they wanted immediate feedback on the answers they provided, and if they did, they were provided

with a summary of the findings from the DAWBA assessment and they were enabled to search for relevant books, websites, helplines and local services. Feedback from youth*in*mind always talks about probabilities rather than certainties, for example saying that on the basis of the information provided, the individual is in the average, slightly raised, high or very high probability bands for a particular type of difficulty. If the computer banding seems out of keeping with the parents' own sense of whether their child has a problem, they are told that the fault may well lie with the computer rather than with their intuitions. While some parents find the automated feedback and linked information adequate for their needs, others want to turn to specialist advice. This is not discouraged in any way; indeed, the website's directory of services makes it easier to find an appropriate source of advice. There is no evidence this leads to inappropriate demands on services.

For the analyses required by the DH remit, youth*in*mind teamed up with two partner organizations in order to study two samples at very different risks of mental health problems: a high-risk sample and a low-risk sample.

- 1) The **high-risk** sample was obtained by collaboration with **HemiHelp**, a voluntary organization for childhood hemiplegia, also known as hemiplegic cerebral palsy. This is a condition involving weakness or stiffness of just one side of the body (either left or right) that dates back to pregnancy, birth or childhood. It affects up to one in a thousand individuals and is usually due to a malformation, stroke, infection or accident. Previous research has shown that around 50% of children and teenagers with hemiplegia have a mental health disorder, often interfering more markedly with the individual's (and family's) quality of life than the physical disability itself (Goodman and Graham, 1996; Goodman and Yude, 2000). A psychiatric disorder rate of 50% is particularly high when you take into account the fact that hemiplegia effectively strikes at random, so there is no association with social deprivation, economic disadvantage or family difficulties. It is also relevant that most children with hemiplegia are of normal intelligence and that their physical disability is relatively mild. The committee of HemiHelp (including parents of children with hemiplegia, and young people with hemiplegia themselves) enthusiastically welcomed youth*in*mind's offer to make the online DAWBA and immediate feedback available without charge to any parent in HemiHelp who wished to use it. This opportunity was brought to the attention of roughly 2,000 parents by emailing all members, as well as by publishing details in HemiHelp's newsletter and on its website. All HemiHelp parents who completed the online DAWBA agreed to the information they provided being anonymized and used for research purposes.
- 2) The **low-risk** sample was obtained in collaboration with the National Centre for Social Research (NatCen), the survey organization that carries out the annual Health Survey for England (**HSE**) on behalf of the NHS Information Centre. The 2009 HSE asked the parents of a representative sample of children and teenagers about various aspects of one or two of their children (but did not ask about these children's mental health). Representative samples of this type have been shown

by previous national surveys to have roughly a 10% rate of significant mental health disorders. This is clearly much lower than the 50% rate in childhood hemiplegia, but the fact that 10% is considered a low rate may surprise those who are more familiar with physical rather than mental health – such a rate would be shockingly high if it applied to epilepsy, diabetes or leukaemia. It is an under-appreciated fact that emotional, behavioural and concentration difficulties are far the commonest severely impairing health disorders affecting children and adolescents. Describing 10% as a relatively low rate is clearly not to argue that this rate is somehow acceptable – it clearly isn't. Following the completion of the 2009 Health Survey for England, NatCen – with the consent of the NHS Information Centre and the NatCen Ethics Committee – agreed to send out invitation letters to around 3,500 parents who had participated in the 2009 survey and who had already agreed in principle to being contacted again.

### **Data collection and analysis**

Online DAWBAs were administered to the HemiHelp sample from January until June 2010, and to the HSE sample from April until July 2010. Data analyses were carried out in the late summer and autumn of 2010, with an initial draft of this report first being submitted in September 2010, and subsequently being revised in the light of very helpful feedback from DH.

### **The main findings on the early years DAWBA**

Three different lines of evidence demonstrated that the early years DAWBA for 2-4 year olds was comparable to the standard school-age DAWBA:

- 1) The early years DAWBA and the standard DAWBA were comparably good at discriminating between the high risk (HemiHelp) and low risk (HSE) groups. For technical details, see Section C.1 of the Technical Report beginning on page 22.
- 2) The two versions of the DAWBA were comparably good at predicting health service use for mental health problems (details in C.2, page24).
- 3) The internal coherence of the two versions was comparable (details in C.3, page 24)

Though all these conclusions would be more secure with larger sample sizes and narrower confidence intervals, they collectively make a convincing case. Since the standard DAWBA has been a suitable measure for previous nationwide surveys of the mental health of 5-16 year olds, it seems reasonable to conclude that the early years DAWBA would be equally suitable for future nationwide surveys that include 2-4 year olds.

### **Policy implications of the finding on the early years DAWBA**

The nationwide surveys that have provided such useful information to policy makers on the mental health of schoolchildren can now be extended to 2-4 year olds – there is now a suitable measure.



It is worth reflecting on what it would take to acquire as much information on 2-4 year olds as we already possess for 5-16 year olds. If we combine the two nationwide surveys of 5-16 year olds (Meltzer et al, 2000; Green et al, 2005), these provide detailed information on the mental health of a representative sample of around 1,600 individuals in each year band. To collect an equivalent number of 2, 3 and 4 year olds would require a survey of around 4,800 preschool children. Even in an age of austerity, it would be a real shame to drop below 1,000 per year band (i.e. 3,000 in total) since the preschool years are a time when children are developing quickly. This makes it highly desirable to have enough in each year band to analyse that band separately. For instance, if it were necessary to combine preschool children into a single group in order to have sufficient statistical power for analysis, this would prevent looking for important differences within this period. Change is particularly rapid in the early years and while it may be sensible to combine, say, 9 and 10 year olds, it would be unwise to have to combine 2 and 3 year olds. The need to sample evenly across a year band is also particularly relevant to preschool children. It would probably not matter much if nearly all 9 year olds in a 5-16 sample were “almost 10 year olds”. It would matter far more if nearly all the 2 year olds in an early years sample were aged 35 months rather than being evenly spread between 24 and 35 months.

### **The main findings on the advantages and disadvantages on online assessment**

In line with previous results from child and adolescent mental health surveys involving a trained interviewer visiting the family at home, the rate of significant mental health disorders was high (53%) in the group with hemiplegia (details in A.3, page 13) and much lower (10%) in the HSE group drawn from the general population (details in B.7, page 20). This provides considerable reassurance that the online DAWBA was doing its job satisfactorily. So too does the fact that the online assessment replicated previous findings on risk factors and service provision (details in B.8, page 21). However, before concluding that online assessments are clearly the way of the future, it is vital to register the proviso that online assessments may arrive at the right answer in the “wrong way”, and therefore not be perceived as trustworthy.

The key issues are participation rates and selection bias. Most experts attach more weight to a survey’s findings if that survey has a high participation rate (i.e. if around 70% of those invited to take part do take part), and if all sections of the population are equally likely to take part. The online survey methods adopted in this study do not come out particularly well by these criteria.

A previous online survey of child mental health in Norway achieved a participation rate of around 45% in a low-risk sample without needing any financial incentives – not great as judged against a 70% benchmark, but no worse than a comparable Norwegian survey using interviewers (Heiervang and Goodman, 2010). Parental participation in a low-risk sample was substantially worse in Britain than Norway: 14% with one reminder, rising to 28% with a financial incentive and two reminders (details in B.2, page 14; also see Figure 1 on page 15). Since the Norwegian and British surveys was very similar in design, it is tempting to suppose that the difference in response rates between Britain and Norway reflects national differences in social structure and social capital. For example, income

inequality is lower in Norway – with a Gini coefficient of 0.251 in 2000 making it one of the most equal countries in the world, as compared to the UK’s Gini coefficient of 0.345 in 2000. Norwegian parents also work shorter hours on average. Perhaps the greater Norwegian participation reflects some combination of better internet access, fewer time pressures and a greater sense of civic inclusion.

A counter-argument to those who criticise a study on the grounds of a low participation rate is that the key issue is generally not response rate but representativeness. A child mental health survey that obtained an excellent 90% response rate would be of little use if the missing 10% was exactly the 10% of the population with a diagnosable mental health disorder. Equally a survey that approached a million people and obtained a “disastrous” response rate of 1% by enrolling 10,000 participants would be absolutely fine in the (unlikely) event that the 1% was a random subset of the general population.

So was recruitment into the online survey equivalent for all sections of the population? Definitely not. There was substantial variation in recruitment rates according to several socio-demographic characteristics (see bar charts on pages 31-34). After adjusting for overlap between these characteristics, there were four independent predictors of families with a lower participation rate: rented accommodation; a more deprived neighbourhood; having only one biological parent in the household (i.e. single parent or reconstituted family); and Asian ethnicity (details in B.5, page 17) At face value, this is a serious problem since it undermines representativeness. The counter-argument is that previous studies (Wolke et al, 2009; Heiervang and Goodman, 2010) have shown that survey findings are often remarkably resistant to this sort of bias – a conclusion that seems to be true in this case too (details in B.7 and B.8, page 20 onwards). Were surveys not resistant to unrepresentativeness, it would be hard to explain why the online survey’s findings on prevalence, risk factors and service use generally matched the findings from traditional interviewer-based surveys with high participation rates.

A definite advantage of face-to-face interviews is that respondents will normally complete the entire interview. By contrast, about a third of respondents to the online DAWBA logged off without having completed the entire interview, though there was evidence that this was mostly a case of respondents not completing sections that did not apply to their child (details in B.6, page 18).

In Norway, switching from interviewers to online assessment reduced the cost of a survey by about 75%. In Britain, the saving would be far more modest (around 30%) if financial incentives and repeated reminders were used to drive up response rates (details in B.4, page 16).

### **Policy implications of the finding on the advantages and disadvantages of online assessment**

If a government wanted to obtain reasonably accurate information about rates of disorder, risk factors and service use, it could probably do so relatively cheaply by inviting a large number of parents recruited from email lists to take part in an online survey without offering any financial incentive. The participation rate would be low but the unit cost per

achieved interview would also be low. Unequal participation of different social groups would occur, but probably wouldn't seriously distort the findings. However, even if this approach generated cheap and fairly accurate findings, there is a real risk that it would not generate credible findings. Such surveys will probably be regarded as second rate by traditional epidemiologists and statisticians – and perhaps even dismissed as an underhand way of massaging government figures by excluding disadvantaged groups.

Even if it is hard to imagine national child and adolescent mental health surveys switching entirely to online assessments in the near future, it is possible to imagine the costs of traditional surveys being reduced by switching to multiple interview methods. If face-to-face and online DAWBAs perform comparably, why not offer respondents a choice? Some respondents will prefer to have an interviewer visit them at home, but others will prefer to respond online (perhaps because this allows them to complete the interview in several instalments when their children are asleep). Other respondents may prefer telephone interviews – and the telephone can also be used to fill in gaps in online interviews. Since online and telephone interviews cost less than sending an interviewer to the family home, everyone wins: respondents choose the method that suits them best; participation rates are likely to increase since families who do not want to be interviewed in person may be willing to complete the interview online; and costs are likely to drop since the cost of an online DAWBA (roughly £5) is much cheaper than the cost of a face-to-face DAWBA (roughly £50). The most convincing way to evaluate these potential advantages would be to carry out a randomised trial. For example, a survey of mental health that will be approaching 10,000 families could randomly assign each family to one of two groups: those who would only be offered a face-to-face interview; and those who would be offered a choice of interview modalities – comparing the average cost and recruitment rate in the two groups.

### **Findings on offering immediate computerized feedback to respondents**

Once they had finished the online assessment, parents from the HemiHelp and HSE samples were offered the chance to get immediate feedback on their answers. The feedback screens included a button that allowed the respondents to tell us what they thought of the feedback. As is often the case with consumer satisfaction surveys, the response rate was low, limiting the extent to which we can generalize from the answers. It is striking, however, that all 20 of the parents who responded thought that it was a good idea to offer parents feedback – including the 6 parents who found the feedback disappointing or very disappointing (details in section D, page 25). There is much more work to be done to improve the quality of feedback and to establish the relative balance of advantages and disadvantages of different sorts of feedback. For the moment, this small pilot study is an insufficient basis for Ethics Committees either mandating or prohibiting such feedback.

# Technical Report

## Contents

<b>A) Analyses based exclusively on the HemiHelp sample</b>		Page
	<i>A.1) HemiHelp: a sample with a high rate of psychiatric disorders</i>	12
	<i>A.2) Recruitment and representativeness of participants</i>	12
	<i>A.3) Psychiatric disorder in the HemiHelp sample</i>	13
<b>B) Analyses based exclusively on the HSE sample</b>		
	<i>B.1) HSE: a sample with a relatively low rate of psychiatric disorders</i>	14
	<i>B.2) Recruitment of HSE parents to the online mental health survey</i>	14
	<i>B.3) Both reminders and incentives increase participation rates</i>	16
	<i>B.4) Is the cost of pushing up the response rate worth it?</i>	16
	<i>B.5) Representativeness</i>	17
	<i>B.6) Completeness</i>	18
	<i>B.7) Prevalence estimates</i>	20
	<i>B.8) Patterns of association</i>	21
<b>C) Validation of the early years DAWBA (HemiHelp and HSE data)</b>		
	<i>C.1) Validation by comparing rates of disorder from high and low risk samples</i>	22
	<i>C.2) Validation by prediction to service use</i>	24
	<i>C.3) Validation by examining the coherence of SDQ and DAWBA measures of disorder</i>	24
<b>D) Parents views about immediate feedback (HemiHelp and HSE data)</b>		25

### A) Analyses based exclusively on the HemiHelp sample

#### *A.1) HemiHelp: a sample with a high rate of psychiatric disorders*

As described in the narrative report, HemiHelp is a voluntary organization for childhood hemiplegia, also known as hemiplegic cerebral palsy. It is a condition involving weakness or stiffness of just one side of the body (either left or right) that dates back to pregnancy, birth or childhood. This physical disability affects up to one in a thousand individuals and is usually due to a malformation, stroke, infection or accident. It is known to be associated with a high rate of psychiatric disorder (Goodman and Graham, 1996; Goodman and Yude, 2000).

#### *A.2) Recruitment and representativeness of participants*

In early 2010, HemiHelp sent an email to roughly 2,000 families with hemiplegic children aged between 2 and 17 years, inviting them to log on to a special website if they were interested in completing the DAWBA online and getting instant automated feedback on their child. Parents completed at least the initial SDQ on 143 individuals

with hemiplegia (39 aged 2-4, 104 aged 5-17). HemiHelp is an organization whose membership may itself be an unrepresentative sample of all children with hemiplegia, and any unrepresentativeness might have been compounded by selective participation in the online survey. With a participation rate of around 7%, it is plausible but by no means certain that the survey particularly attracted parents who were actively concerned about their child's mental health at the time they received the email from HemiHelp. We were not in a position to examine selection bias directly since we had not sought permission to access HemiHelp's own records to identify the characteristics of non-responders. However, there are indications that the children of HemiHelp members who did participate in the online survey were similar to representative samples of children with hemiplegia who have been studied previously. In this sample as in previous representative samples, there were more affected boys than girls (60% boys in this sample) and the right side of the body was more often affected than the left side (58% right-sided in this sample). Did this similarity extend to psychiatric disorder? That question is addressed in the following section.

*A.3) Psychiatric disorder in the HemiHelp sample*

It is instructive to compare the rates and types of psychiatric disorder found in the HemiHelp online survey with the rates and types of psychiatric disorder found in previous studies of a representative sample of children with hemiplegia, the London Hemiplegia Register (Goodman and Yude, 1996). The comparison was restricted to 5-17 year olds since this is the age range that has been the main focus of previous studies.

Of the 96 children in the HemiHelp survey whose parents provided enough information to make a diagnostic rating, 53% had one or more ICD-10 psychiatric disorders. For comparison, the rate of psychiatric disorder in the previous representative sample was variously estimated at 55% (based on parent questionnaire alone) and 61% (based on a detailed multi-informant psychiatric assessment). Table A.3.1 compares the rates of the main groups of disorders, demonstrating a close similarity for most disorders. The higher rate of autistic spectrum disorders (ASDs) in the online survey (10% v. 4%) may reflect selective participation of their parents, but it is just as plausible that it reflects the fact that we now have a broader concept of ASDs than we did 20 years ago, along with improved assessments.

Table A.3.1

	Prevalence of disorder (%)	
	London Hemiplegia Register, multi-informant psychiatric assessment N=149	HemiHelp online survey N=96
Any disorder	61%	53%
Emotional disorder	25%	25%
Behavioural disorder	24%	28%
Hyperkinetic disorder	10%	7%
Autistic spectrum disorder	4%	10%

Though the numbers of children are too small to warrant showing rates for each individual disorder, it is worth adding that the online survey also replicated the established finding that hemiplegia does not increase the prevalence of all symptoms and disorders to the same extent: worries and fears are much increased, but depression, obsessions and compulsions are not; irritability, oppositionality and defiance are much commoner, but proactive aggressive and antisocial behaviours are not (Goodman and Yude, 2000).

In conclusion, the online assessment of the HemiHelp sample demonstrates the enormous potential for online “convenience” samples to generate “good enough” results very economically. Sending out a group email generated findings that were remarkably similar to the results of the much more labour-intensive London Hemiplegia Register sample. Assembling that representative sample some 20 years ago occupied about 50% of the time of a full-time research fellow for 3 years. Perhaps we should not be too surprised by the usefulness of convenience samples, even if they do have a lower scientific prestige than representative samples. It is worth remembering that convenience samples have proved useful in many previous sorts of research, e.g. for developmental psychology studies of mothers and children recruited via adverts in local papers, or for “behavioural phenotype” studies of children with specific genetic disorders recruited via the relevant parents’ groups.

## **B) Analyses based exclusively on the HSE sample**

### *B1) HSE: a sample with a relatively low rate of psychiatric disorders*

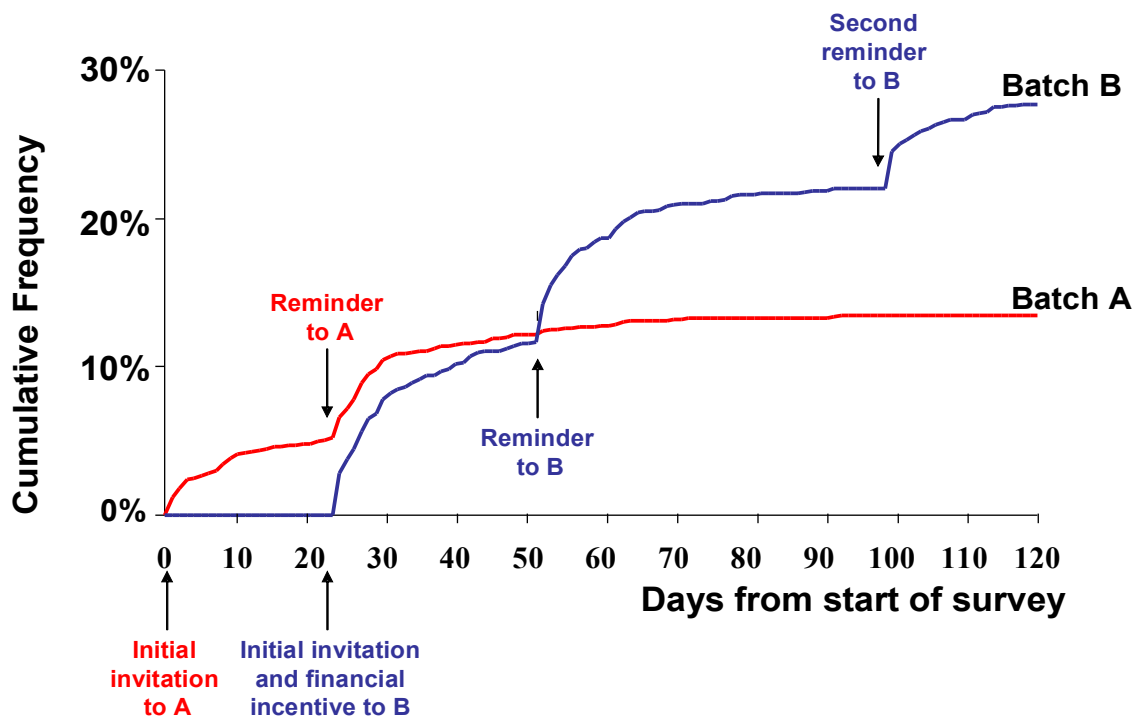
The National Centre for Social Research (NatCen) is a leading survey organization whose work includes the annual Health Survey for England (HSE), carrying this out on behalf of the NHS Information Centre. The 2009 HSE asked parents about various aspects of one or two of their children (but did not ask about these children’s mental health). The survey also collected socio-demographic information about the parents and family as a whole: family composition, household income, housing tenure (owned v. rented), social class, parental education, ethnicity, area of residence (urban v. rural), region, and area deprivation (Index of Multiple Deprivation).

### *B2) Recruitment of HSE parents to the online mental health survey*

These were potentially 3,500 young people aged between 2 and 15 whose parents had participated in the 2009 HSE and who had agreed that they could be contacted again. These parents were invited to log onto a specific website, using the ID number and password provided. The mail-out of invitation letters was piloted on 100 families, revealing one bug in the system that was rapidly mended. The remaining invitation letters were mailed out in two batches, separated by three weeks: Batch A (N=1,700) was sent out in late April 2010 and Batch B (N=1,752) was sent out in mid-May. The division of families into Batches A and B was not based on any systematic selection criteria, but simply followed the order of an internal ID number created for the 2009 HSE survey.

In general, most people invited to take part in online surveys respond rapidly or not at all (unless they are sent reminders). By early May, it was clear that the response rate in Batch A was much poorer than anticipated – around 5% as opposed to the 15-25% predicted from the response to the initial invitation in the comparable “Children in Bergen” study in Norway. In consultation with DH and NatCen, we decided as a matter of urgency to alter the planned protocol so that Batch B received a financial incentive with their invitation letter (having received permission for this modification to the protocol from NatCen’s Ethics Committee). It was decided not to offer an incentive to Batch A retrospectively since although this would probably have led to more total responses, it was only by retaining a no-incentive comparison group that we could establish for future reference what effect the incentive was having on participation.

**Figure 1. Participation in online survey following mail-out of invitation letter and reminder letters**



In the light of NatCen’s previous experience of which incentives work best, the chosen incentive was a £5 voucher sent “up front”, i.e. as an unconditional token of appreciation. This has previously proved to be a better motivator than promising a larger reward that will not be sent out until after the recipient has completed the interview. It may strike readers as counter-intuitive that unconditional payments work better, but it seems that a substantial number of respondents who would not have been sufficiently motivated by the promise of £5 do nevertheless respond to the gift of £5 – to reciprocate the generosity or out of a sense of obligation. Batch A were sent a reminder letter if they had not responded after 3 weeks. Batch B were sent up to two such reminders if they still had not responded after 4 and 11 weeks. The results are summarised in Figure 1, where

participation is defined in terms of the parent having logged on to the website and completed the compulsory initial Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001). The question of how often respondents complete all the DAWBA sections is postponed until section B.6 of this report (on page 18).

### *B.3) Both reminders and incentives increase participation rates*

As Figure 1 shows, the first reminder roughly doubled the response rate for Batch A and Batch B. The second reminder (sent just to Batch B) was only about half as effective, providing evidence for diminishing returns beyond a certain point.

Figure 1 also demonstrates the value of an incentive – this is evident from the fact that the initial invitation generated roughly twice as many responses from Batch B as compared with Batch A (and the first reminder letter also generated more responses from Batch B than Batch A). While Figure 1 tells a straightforward story that is easy to understand intuitively, there is one caveat. The division into batches was not done by random assignment as occurs in a clinical trial – so it is important to consider the possibility that Batch B just happened to contain far more “easy to recruit” families than Batch A. As described in a subsequent section (B.5, page 17), several socio-demographic characteristics had a major influence on recruitment and Batches A and B differed to a small but statistically significant extent on several of these characteristics. Taking these differences into account, it is possible to use weighting via propensity scores to estimate that if incentives and second reminders had been allocated at random (i.e. independent of batch), the predicted participation rate would have been 21.2% for Batch A and 19.8% for batch B. Since these differences were in the direction of making Batch B slightly *harder* to recruit, the visual impression from Figure 1 that incentives made a big difference to recruitment slightly understates the impact of incentives and additional reminders.

### *B.4) Is the cost of pushing up the response rate worth it?*

By comparison with the 14% participation rate in batch A that was achieved with one reminder and no financial incentive, it was possible to achieve a 28% participation rate in Batch B by using a £5 “up front” incentive plus two reminders. Was this worth the effort and cost? .

To obtain indicative costs, consider the following simplified scenario: each invitation or reminder letter costs £1 (in paper, postage and labour costs), each online assessment costs £5, and “up front” incentives are £5. With an incentive, two reminders and a 25% response rate, achieving 1000 responses would involve:

Sending 4,000 invitations (£4,000)  
Sending 4,000 incentives (£20,000)  
Sending about 3,600 first reminders (£3,600)  
Sending about 3,200 second reminders (£3,200)  
Carrying out 1,000 online assessments (£5,000)  
Total cost = £35,800, i.e. £35.80 for each response.



If the cost of obtaining one response in a traditional survey with face-to-face interviewers is taken to be around £50, this does represent a 30% saving. While that is clearly attractive, it is much less than the estimated 75% saving when the Children in Bergen study switched from face-to-face to online interviewing (Heiervang and Goodman, 2010). The key difference is that the response rate was 45% in Bergen without the need for any financial incentive.

Finally, it is worth noting that both friends and critics of online surveys could argue that achieving a response rate of 28% through significant effort and expense is the worst of all possible worlds – still low enough to lack credibility with some researchers, and yet no longer the much cheaper alternative to a traditional interviewer-based surveys. Using lists of email addresses as a basis for emailing invitations and reminders could potentially cut costs considerably, particularly if no financial incentive was offered, a low response rate was acceptable, and the resultant unrepresentativeness was not seen as an insuperable barrier. Future studies could establish whether this sort of approach does generate useful findings at a sufficiently low cost to compensate for unrepresentative participation.

#### *B.5) Representativeness*

Though methodological reviews often criticise studies with low response rates, this is somewhat misleading. The key issue is generally not response rate but representativeness. We therefore need to ask whether the low response rate in the online HSE follow-up led to unrepresentativeness. It was possible to examine this since NatCen had collected socio-demographic information about families when they first participated in the 2009 Health Survey for England. The results are displayed as bar charts at the end of the appendix, showing participation rates for each socio-demographic indicator. As previously noted, participation is defined narrowly as completion of the compulsory initial SDQ.

To summarize the univariate effects shown in the bar charts, there are clear and highly significant linear trends for participation to rise with income ( $p < .001$ ) and social class ( $p < .001$ ), and to fall with area deprivation ( $p < .001$ ). The trend for participation to fall with increasing family size was only marginally significant ( $p = .084$ ). For the binary variables, participation was significantly lower for those living in rented accommodation ( $p < .001$ ), those living with just one biological parent ( $p < .001$ ) and those living in urban areas ( $p < .005$ ). For the categorical variables, there was an overall effect for geographical region ( $p < .001$ ) and ethnicity ( $p < .002$ ). The values of adjusted residuals demonstrated significantly higher participation in the Eastern and South Eastern regions, and significantly lower participation in the North West region and among British Asians.

While many socio-demographic factors predicted participation, these predictor variables were, in many instances, highly correlated with one another, e.g. low household income was closely related to low social class, living in rented accommodation and living in deprived areas. Multivariate logistic regression was therefore carried out to identify the variables that predicted participation even after adjusting for all the other predictors. Data were missing on less than 1% of the sample for all variables except for household income (14% missing). Having dropped household income, forward conditional

regression was carried out with the remaining socio-demographic variables as the independent variables and with completion of the initial SDQ as the dependent variable. Region and ethnicity were treated as categorical variables; the entry criterion was  $p < .02$ . There were four significant independent predictors:

- 1) *Rented housing* – parents who rented their home were less likely to participate than parents who owned their home outright or with a mortgage (Odds Ratio 0.47,  $p < .001$ ).
- 2) *Non-traditional family* – single-parent and reconstituted families were less likely to participate than families where both biological parents were resident in the household (Odds Ratio 0.74,  $p = .006$ )
- 3) *More deprived area* – parents living in the more deprived areas were less likely to participate (Odds Ratio 0.92 per quintile increase,  $p = .013$ ).
- 4) *Ethnicity* influenced participation after adjusting for the previous variables ( $p = .02$ ), with Asian families being significantly less likely to participate (Odds Ratio 0.29 by comparison with White families,  $p = .007$ ).

Based on these four predictors, it was possible to generate a “propensity score” for each individual, i.e. the predicted likelihood of participating based on the relevant socio-demographic factors. The reciprocal of this score was then used to weight the sample in order to adjust for the over- and under-representation of specific socio-demographic groups. This weighting was used in some analyses presented in this report.

#### *B.6) Completeness*

When respondents miss out questions or entire sections, this makes analysis harder. The Norwegian experience was that very few respondents failed to complete the entire DAWBA interview when this was administered by an interviewer, whereas around half left one or more DAWBA sections blank when completing the DAWBA online (Heiervang and Goodman, 2010). This difference can be at least partially explained in terms of social conventions. It generally seems rude to people to put the phone down on someone or terminate a personal interview prematurely, but most people have fewer inhibitions about hopping between TV channels or terminating an internet interview that is getting tedious.

It is impossible to stop an online respondent logging off, and this needs to be recognized in the design of online interviews. It is certainly possible to appeal to respondents’ benevolence and request that they complete everything if possible. Beyond this, years of experience have taught us at youthmind that it is best to give respondents scope to answer what interests them most. The DAWBA begins with the SDQ, a brief screening instrument that is likely to be relevant to everyone. From that point onwards, respondents can choose where to go next from a menu of DAWBA sections with options such as: “Depression”, “Hyperactivity and attention problems” and “More about your child’s strengths and good points.” Parents are likely to be more motivated to complete sections that seem particularly relevant to their child. If they have only limited time or enthusiasm, it is better if they use that to record their concerns rather than forcing them to go through all the sections in a fixed order – with the risk that they will get bored and

log off before describing problem areas. The underlying assumption is that skipped sections would probably have been irrelevant anyway. Of course, it may sometimes happen that parents omit a section because their child has serious problems in this area that they don't want to talk about. However, experience of following up unanswered DAWBA sections during clinical interviews suggests that this is unusual. Thus clinical experience suggests that "not answered" generally turns out to mean "no significant problems in this area" (but clinical experience does not always survive scientific testing).

Once parents have completed the SDQ and reached the DAWBA menu, they are free to answer between zero DAWBA sections and all of them (up to 15 sections for 5-17 year olds; up to 13 sections for 2-4 year olds). The HSE sample completed the SDQ on 725 children, and the breakdown of the number of DAWBA sections is shown in Table B.6.1:

Table B.6.1

Number of DAWBA sections completed	Number of respondents this applied to (%)	Number (%) in this category who were predicted to have a disorder by the SDQ	Number (%) in this category who were assigned a clinical DAWBA diagnosis
None	101 (14%)	1 (1.0%)	[not rateable]*
At least one section, but not all of them (mean number of sections completed=5.6)	156 (21%)	17 (10.9%)	22 (16.7%)
All of them	468 (65%)	21 (4.5%)	30 (6.4%)
Total sample	725 (100%)	39 (5.4%)	52 (7.2%)

\* These individuals were not rateable because no part of the DAWBA had been completed

The 14% of parents who did not complete any sections of the DAWBA were the ones whose children had the lowest rate of probable disorders as judged from the SDQ – less than a quarter of the rate for the sample as a whole. This supports the view that the main reason parents stop after completing the SDQ is because they have no cause for concern and therefore see no reason for choosing any DAWBA section.

The 21% of parents who answer some but not all DAWBA sections have children with double the rate of probable disorders as judged from the SDQ, and also double the rate of subsequent DAWBA diagnoses. This agreement between the SDQ and DAWBA is reassuring as it suggests that partial completion has not undermined the ability of the DAWBA raters to assign the expected proportion of DAWBA diagnoses.

As judged from the SDQ, the 65% of parents who answer all sections have children with a rate of probable disorders slightly below the rate for the sample as a whole, just as their rate of subsequent DAWBA diagnoses is slightly below the rate for the sample as a whole. This is not what would be expected if these parents' motivation to complete every section was a particularly high level of parental concern. Instead, completing every section may reflect parental conscientiousness and altruism, with them providing full

details in order to help the research project as much as possible even if their own children are, on average, doing better than the sample as a whole.

In summary, although it would be more convenient for statistical analysis if all respondents answered all sections of the online DAWBA, the problem is not as serious as it could be. Two-thirds of respondents did answer all sections, and even if analyses were restricted to this group (representing what is sometimes known as “listwise deletion of missing data”), it would probably not have distorted findings much since the group had a level of psychopathology close to the group average. Alternatively, if analyses were carried out on the entire sample without dropping subjects with missing data, there is support for the clinical impression that “no answer” typically reflects “no problem in this area”, and that missing sections do not undermine the ability of the DAWBA to detect the expected proportion of disorders.

### B.7) Prevalence estimates

One important indicator of the value of using the online DAWBA is whether it can generate accurate prevalence estimates for psychiatric disorders. The “correct” DAWBA prevalences can be taken as the rates of disorder in 5-15 year olds established by the 1999 and 2004 British child and adolescent mental health surveys (B-CAMHS), thereby drawing on a sample of almost 18,000 children and adolescents. These rates can be compared with rates derived from the online HSE sample (restricted to 5-15 year olds with sufficient DAWBA information to form the basis for a diagnostic assessment; N=462). In Table B.7.1, the HSE rates are presented with and without weighting to adjust for selective participation in the online survey. When used, this weight adjusts for the under-representation of children from more deprived areas, living in rented housing, coming from non-traditional families, or of Asian ethnicity. The rates shown in Table B.7.1 are restricted to the main categories of disorder – the HSE sample size is too small to justify presenting results for specific disorders one by one.

Table B.7.1

ICD-10 Disorder	Prevalence in 5-15 year olds (%)			
	B-CAMHS		HSE	
	1999 N=10,438	2004 N=7,363	Not weighted N=462	Weighted N=462
Any disorder	9.5	9.6	9.6	10.7
Emotional	4.3	3.5	4.4	4.5
Behavioural	5.3	5.9	3.7	4.8
Hyperkinesis	1.4	1.5	0.7	0.6

Note: The sum of emotional, behavioural and hyperkinetic disorders does not necessarily add up to the rate of “any disorder” since (a) some individuals have more than one type of disorder, and (b) some disorders such as autism or anorexia are included in the total but are not classified as emotional, behavioural or hyperkinetic disorders.

It can be seen from Table B.7.1 that the HSE rates are relatively close to the “correct” 1999 and 2004 rates based on much larger samples. The exception is hyperkinesis, but

this was based on a positive diagnosis in just three HSE individuals and the 95% confidence interval is very wide (0 to 1.7%). It is not surprising that weighting to adjust for selective participation particularly improves the match for behavioural disorders since these disorders are commoner among children from less affluent backgrounds, from non-traditional families, and from poorer areas – and hence the prevalence of such children is particularly likely to be under-estimated in online surveys that selectively under-sample these groups.

The HSE sample was specifically chosen because its socio-demographic characteristics had already been studied, permitting adjustment for uneven participation. This is not true for most online surveys, where the researchers only know about those people who take part and not about those who do not take part – making it harder to adjust for non-participation. However, the current findings suggest that lack of adjustment would not be a disaster as far as prevalence estimates are concerned. The unweighted prevalence estimates may not have been exactly right but they are arguably “good enough” for service planning, for which ‘ball park’ figures are often all that is required.

*B.8) Patterns of association*

Surveys are not just about estimating prevalence, important though this is for planning services and detecting secular trends. The most important findings of surveys are often the pattern of associations that they reveal, for example the associations between risk factors and disorders (relevant to aetiology and prevention), between different disorders (comorbidity), and between disorders and outcome (natural history, prognosis). Previous studies have shown that selective recruitment is less likely to distort patterns of association than estimates of prevalence (Wolke et al., 2009; Heiervang and Goodman, 2010). To see if patterns of association were preserved in the present study, we examined whether it was possible to replicate the well-established associations between having at least one type of psychiatric disorder and four risk or outcome factors selected a priori. The results are summarized in the following tables:

Table B.8.1 *Learning difficulties*

	Presence of any ICD-10 disorder
No learning difficulties	7% (25/369)
Learning difficulties	35% (7/20)

p<.001

Table B.8.2 *Adverse life events in the previous 12 months*

	Presence of any ICD-10 disorder
None	6% (16/268)
One	7% (7/97)
Two or more	38% (9/24)

p<.001

Table B.8.3 *Financial difficulties in the family*

	Presence of any ICD-10 disorder
No	7% (18/260)
A little	8% (8/105)
A lot	27% (6/22)

p<.004

Table B.8.4 *Contact with child and adolescent mental health services (CAMHS)*

	Has received help or advice from CAMHS
No ICD-10 disorder	2% (6/350)
One or more ICD-10 disorders	47% (15/32)

p<.001

Though far from exhaustive, these unweighted analyses do confirm that online surveys can detect meaningful patterns of association.

### **C) Validation of the early years DAWBA**

The strategy used to validate the early years DAWBA relied on combining and contrasting the two samples described earlier in this report, namely the HemiHelp sample with its high rate (53%) of psychiatric disorders and the HSE sample with its relatively low (10%) rate of psychiatric disorders. As reported in sections A.3 and B.7 of this technical report, the online samples had rates of disorder that were very similar to those found in previous studies of representative samples of children with hemiplegia or children from the general population. This makes it very unlikely that the validation findings reported here were distorted by collecting the information online rather than using traditional face-to-face interviews. Equally, it seems very implausible that the validation findings have been significantly distorted by the lower participation rates accompanying online assessments.

In both the HemiHelp and HSE samples, data were collected using the standard DAWBA for 5-17 year olds, and using the early years DAWBA for 2-4 year olds. Contrasting and combining these two samples forms the basis for three different methods for establishing whether the early years DAWBA is a valid measure of child mental health.

#### *C.1) Validation by comparing rates of disorder from high and low risk samples*

If we had a true “gold standard” measure in child and adolescent psychiatry, we could validate the early years DAWBA by examining how well its diagnoses compared with the gold standard diagnoses. Unfortunately, we have no such gold standard assessment – all we have are assessments that are mediocre to varying degrees (what has been dubbed a “lead standard”). What this means is that we cannot unequivocally validate the early years DAWBA by looking for agreement with some widely recognized measure of preschool psychopathology such as the PAPA (Egger et al, 2006). If the PAPA and DAWBA agreed well, this could simply represent imitation, fashion or a folie à deux. In thirty years time, experts may well look back on the PAPA and the early years DAWBA and see that they were both seriously wrong, but that they agreed well with one another

because they were wrong in the same sort of way. The other way round, if the early years DAWBA and PAPA disagree, there is no straightforward way of telling which is right.

Despite these reservations, it may one day be worth carrying out a head-to-head comparison of the PAPA and the early years DAWBA. However, doing this will have its own problems because comparing a moderately long diagnostic assessment (the DAWBA) with an even longer diagnostic assessment (the PAPA) is not as easy as it sounds. Most respondents are simply not up to completing both measures in rapid succession – the repetition would be too boring or exhausting for them. Furthermore, respondents rapidly learn that they can cut a tedious interview short by answering “no” to screening questions, thereby avoiding all the follow-on questions that would have been asked if they had said “yes”. This leads to far fewer positive answers in the second interview, a process known as “attenuation”. It is possible to avoid attenuation by waiting around 6 months before administering a second interview – but by this time the child may genuinely have changed so it is no longer reasonable to expect the two interviews to come up with the same result.

Given the lack of a gold standard, and given the phenomenon of attenuation, how can the early years DAWBA be validated? A good approach is to examine how well the DAWBA can distinguish between the HSE low-risk sample and the high-risk HemiHelp sample. Whatever other doubts experts may harbour about diagnostic assessments and their validity, well informed clinicians and researchers can hardly doubt that children with hemiplegia have far more problems with emotions, behaviour, concentration and relationships than children from the general population. This provides child psychiatry’s equivalent of the fixed point from which we can move the world!

Statistically, the approach translates into using logistic regression to see how much commoner it is for the DAWBA assessment to lead raters to diagnose ICD-10 diagnoses in 2-4 year olds with hemiplegia than in 2-4 year olds from the normal population. As a guide to interpreting an Odds Ratio, a ratio of 1 would mean that the rate (technically the “odds”) of a diagnosis was the same in the two groups, while a ratio of 5 would mean that the rate was five times as high in hemiplegia as the normal population. How high would the ratio need to be to demonstrate that the early years DAWBA was “good enough”? The simplest answer is that if the early years DAWBA can do more or less as well as the standard DAWBA, that would be very reassuring since the standard DAWBA has a well established track record for providing good enough data for the national child and adolescent mental health surveys carried out over 10 years. As shown in Table C.1.1, Odds Ratios for the early years and standard DAWBAs are roughly comparable: the difference between an Odds Ratio of 27.7 for the early years DAWBA and 10.5 for the standard DAWBA may look marked, but it is not statistically significant and the 95% confidence intervals overlap.

Table C.1.1

	Age 2-4	Age 5-17
Number of valid subjects* from HSE	98	502
Number of valid subjects* from HemiHelp	30	96

Odds Ratio of at least one ICD-10 diagnosis (HemiHelp v. HSE)	27.7	10.5
95% confidence interval around the Odds Ratio	7.1 – 107.4	6.4 – 17.2

\* Valid subjects are those with sufficient DAWBA data to permit a diagnostic evaluation.

### *C2) Validation by prediction to service use*

A second way of assessing the validity of the early years DAWBA is to examine how well the presence or absence of at least one ICD-10 DAWBA diagnosis predicts whether the family has sought advice or help for mental health problems from primary care, paediatrics or Child and Adolescent Mental Health Services. This was done by examining Odds Ratios in the combined sample of children from HSE and HemiHelp. It is important to note that the diagnostic ratings were made from the DAWBA sections on psychological problems while remaining blind to the answers to the DAWBA questions on service use. As shown below in Table C.2.1, the Odds Ratios for the early years and standard DAWBAs are again comparable, with overlapping confidence intervals and no statistically significant difference.

Table C.2.1

	Age 2-4	Age 5-17
Number of valid subjects*	113	473
Odds Ratio of having sought help from the health service for mental health problems** (ICD-10 diagnosis v. no ICD-10 diagnosis)	23.7	19.1
95% confidence interval around the Odds Ratio	4.9 – 114.5	10.4 – 34.9

\* Valid subjects are those from HemiHelp or HSE with (a) sufficient DAWBA data to permit a diagnostic evaluation and (b) sufficient DAWBA data to identify whether or not the family have sought advice or help for mental health problems from primary care, paediatrics or CAMHS.

\*\* The Odds Ratios are adjusted to take account of which sample the children are drawn from (i.e. HemiHelp or HSE). This is to ensure that a DAWBA diagnosis is not simply predicting service contact because it is a marker for which children have hemiplegia.

### *C3) Validation by examining the coherence of SDQ and DAWBA measures of disorder*

A third approach to assessing the validity of the early years DAWBA takes advantage of the fact that the DAWBA generates three different measures of whether or not a child has a disorder:

- 1) The DAWBA begins with the SDQ, and the symptom and impact scores from the SDQ can be used to predict whether or not the child has a probable disorder (Goodman et al, 2000b). This is referred to subsequently as “SDQ caseness”.
- 2) The answers to structured DAWBA questions about psychological problems are fed into a computerised algorithm that generates up to 6 probability bands that can be collapsed into a binary rating of disorder or not (Goodman, A et al, 2010). This binary rating is referred to subsequently as “DAWBA computer caseness”.



- 3) A trained clinical rater reviews all available evidence from the DAWBA, including the answers to structured questions, the transcripts of respondents' open-ended descriptions of problems, and the computer predictions. On this basis, the clinical rater decides whether to assign diagnoses, and if so, which ones. These are referred to subsequently as "DAWBA clinical ratings". It is precisely these clinical ratings that are used in all the diagnostic analyses described in this report.

Given that the DAWBA involves these three different measures of whether or not a child has a disorder, it is possible to examine the coherence of the different measures. If the early years DAWBA is working well, SDQ caseness and DAWBA computer caseness should be highly predictive of the DAWBA clinical rating. Once again, this was done by examining Odds Ratios in the combined sample of children from HemiHelp and HSE, using the findings for the standard DAWBA as the benchmark. As shown below in Table C.3.1, SDQ caseness and DAWBA computer caseness are highly predictive of a clinician rated diagnosis, with equivalence between the standard and early years DAWBA in terms of overlapping confidence intervals and no statistically significant differences.

Table C.3.1

Predicting from:	Odds Ratio of having an ICD-10 DAWBA clinical rating (95% confidence interval)	
	Age 2-4 N=128*	Age 5 -17 N=598*
SDQ caseness v. not	30.1 (7.6 – 119.6)	28.6 (16.4 – 49.9)
DAWBA computer caseness v. not	48.4 (8.9 – 263.1)	68.0 (36.1 – 128.0)

\* Valid subjects are those from HemiHelp or HSE with sufficient DAWBA data to permit a diagnostic evaluation.

In summary, three different lines of evidence converge to suggest that the early years DAWBA is comparably valid to the standard DAWBA.

#### **D) Parents views about immediate feedback (HemiHelp and HSE data)**

The HemiHelp and HSE respondents were offered the opportunity to receive immediate feedback on the information that they had provided on their children, leading on to the possibility of searching for potentially relevant books, websites, helplines and services. No one was forced to see feedback if they did not want to. There were several reasons for offering this feedback option: in order to motivate respondents; because it is part of youth*in*mind's philosophy that respondents should be free to choose to get feedback on their answers; and because most surveys of physical health (but few if any surveys of child mental health) are required to provide such feedback by Ethics Committees. There are evidence-based treatments for many child mental health problems, and the DAWBA is able to identify children at risk of these problems – i.e. children who might benefit

from more detailed assessment to see if they warrant treatment. Despite all these good reasons for providing feedback, there is an established tradition of offering no feedback from child mental health surveys, including the national child and adolescent mental health surveys. Why is this? There are several powerful arguments in favour of withholding feedback. Firstly, the interviewers who administer traditional child mental health surveys are not trained clinicians and they are not in a position to do anything more than have fact sheets on how to access services – fact sheets that can then be offered to respondents who express concern. Secondly, there is the danger that providing feedback may do more harm than good, whether by raising unwarranted anxiety or by providing false reassurance. Thirdly, there are medico-legal issues about liability for any feedback (though this needs to be set against the medico-legal issues raised by not providing feedback for distressing and potentially treatable problems).

While there is clearly merit in the view that social survey interviewers lack the training to provide feedback, youthinmind is in a better position. We have 10 years experience of providing instant feedback on online SDQs (on [www.youthinmind.info](http://www.youthinmind.info)) and this experience underpins the system for providing feedback after respondents have completed the online DAWBA. On this basis, youthinmind concluded that the likely advantages of providing automated feedback on the online DAWBA outweighed the disadvantages. Were we right? There was no way of knowing whether this feedback would be welcome and helpful without trying it out and asking respondents what they thought. The feedback screens seen by the HemiHelp and HSE respondents included a “Tell us!” button that administered a brief consumer satisfaction questionnaire. As is often the case with consumer satisfaction surveys, the response rate was low – only 2% of respondents (20 out of 893) answered these questions: 8 from HemiHelp and 12 from HSE. Findings based on such a small and potentially unrepresentative sample obviously need to be interpreted with caution. However, even this small survey provides the best guide that we currently have as to whether future DH-funded surveys should offer parents the option of feedback.

The following tables show the consumer satisfaction questions and the responses they elicited, with the “typical” responses shown in bold (choosing the median response for unimodal distributions and the modal responses for multimodal distributions):

Table D.1

In general, do you think it is a good idea to offer parents the chance to see an immediate report on information they have provided about their children?

No	0%	0/20
<b>Yes</b>	100%	20/20

Table D.2

Overall, would you describe the report and information you have just seen as:

Very disappointing	10%	2/20
Disappointing	20%	4/20
<b>Fair</b>	25%	5/20
Good	25%	5/20

Very good	20%	4/20
-----------	-----	------

How well do the following descriptions apply to the report and information?

Table D.3a *Accurate*

Not at all	5%	1/19
A little	26%	5/19
<b>A medium amount</b>	26%	5/19
A great deal	42%	8/19

Table D.3b *Convenient*

Not at all	5%	1/19
A little	21%	4/19
A medium amount	11%	2/19
<b>A great deal</b>	63%	12/19

Table D.3c *Good coverage*

Not at all	11%	2/19
A little	26%	5/19
<b>A medium amount</b>	21%	4/19
A great deal	42%	8/19

Table D.3d *Helpful*

Not at all	11%	2/19
<b>A little</b>	42%	8/19
A medium amount	11%	2/19
<b>A great deal</b>	36%	7/19

Table D.3e *Reassuring*

Not at all	21%	4/19
<b>A little</b>	32%	6/19
A medium amount	32%	6/19
A great deal	15%	3/19

Table D.3f *Upsetting*

<b>Not at all</b>	63%	12/19
A little	32%	6/19
A medium amount	5%	1/19
A great deal	0%	0/20

Table D.4

Was the amount of information on offer:

Too little	32%	6/19
<b>About right</b>	68%	13/19
Too much	0%	0/19

Table D.5

Did you find the computerized assessment and report unpleasantly impersonal?

<b>No</b>	78%	15/19
A little	11%	2/19
A lot	11%	2/19

It is clear that the feedback still has plenty of room for improvement, and youth*in*mind is working on this. Nevertheless, the results are encouraging. All respondents – including the 30% who found the feedback disappointing or very disappointing – felt that it was a good idea to offer parents the chance to see an immediate report on information they have provided about their children. This strengthens the case for continuing to offer parents the chance to see (improved) feedback. Some respondents found the existing feedback very helpful and no one found it very upsetting. It is naturally a cause for concern that one respondent found the feedback moderately upsetting – hopefully, modifying the feedback in the light of continuing satisfaction surveys and focus groups can help avoid upsetting others in future. Having worked hard to personalize the feedback, the fact that only 22% of the respondents found the computerized assessment and report unpleasantly impersonal is encouraging, though no cause for complacency.

## References

Egger, H. L., Erkanli, A., Keeler, G., Potts, E., Walter, B. K., Angold, A.. (2006) Test-Retest Reliability of the Preschool Age Psychiatric Assessment (PAPA). *Journal of the American Academy of Child & Adolescent Psychiatry*, **45**, 538-549.

Field, F. (2010) *The Foundation Years: Preventing poor children becoming poor adults. The report of the Independent Review on Poverty and Life Chances*. HM Government. <http://povertyreview.independent.gov.uk/media/20254/poverty-report.pdf>

Goodman, A., Heiervang, E., Collishaw, S., Goodman, R. (2010) The 'DAWBA bands' as an ordered categorical measure of child mental health: description and validation in British and Norwegian samples. *Social Psychiatry and Psychiatric Epidemiology*, in press.

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.

Goodman, R., Ford, T., Richards, H., Gatward, R., Meltzer, H. (2000a) The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, **41**, 645-655.

Goodman, R., Graham, P. (1996) Psychiatric problems in children with hemiplegia: cross sectional epidemiological survey, *British Medical Journal*, **312**, 1065-1069.

Goodman, R., Yude, C. (1996) Do incomplete ascertainment and recruitment matter? *Developmental Medicine & Child Neurology*, **38**, 156-165.

Goodman, R., Yude, C. (2000) Emotional, behavioural and social consequences. In: Neville, B. & Goodman, R. (eds.) *Congenital Hemiplegia: Clinics in Developmental Medicine, No 150*, London: Mac Keith Press, pp. 166-178.

Goodman, R., Ford, T., Simmons, H., Gatward, R., Meltzer, H. (2000b) Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, **177**, 534-539

Green, H., McGinnity, A., Meltzer, H., *et al.* (2005) *Mental health of children and young people in Great Britain, 2004*. Basingstoke: Palgrave Macmillan

Heiervang, E., Goodman, R. (2010) Advantages and limitations of web-based surveys: Evidence from a child mental health survey. *Social Psychiatry and Psychiatric Epidemiology*, in press.

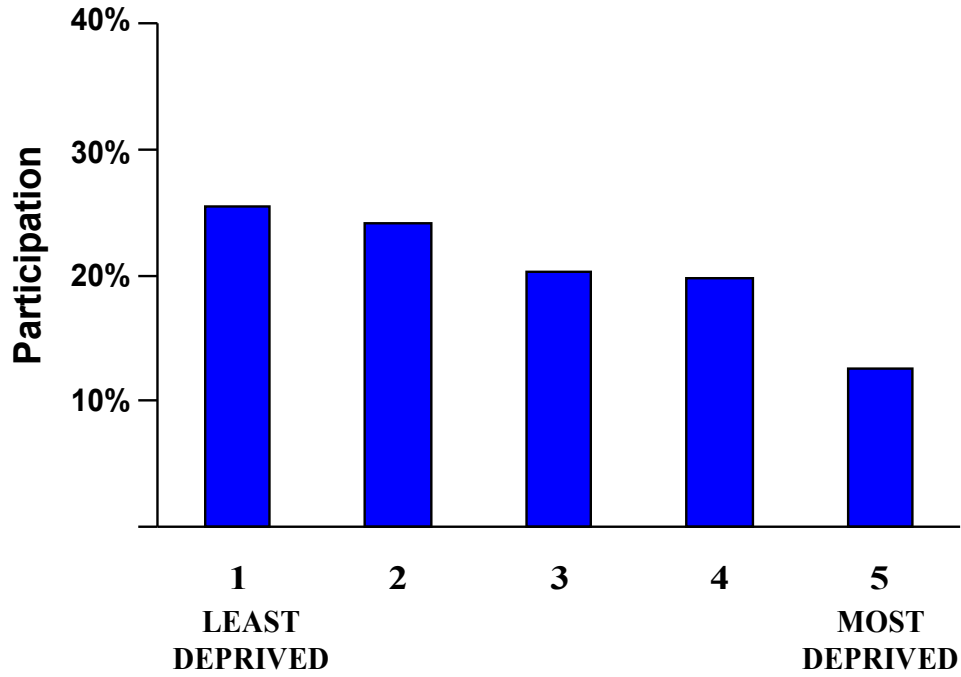
Meltzer, H., Gatward, R., Goodman, R., Ford, T. (2000) *Mental health of children and adolescents in Great Britain*. London: The Stationery Office.

Meltzer H, Gatward R, Corbin T, *et al.* (2003) *The mental health of young people looked after by local authorities in England*. London: The Stationery Office

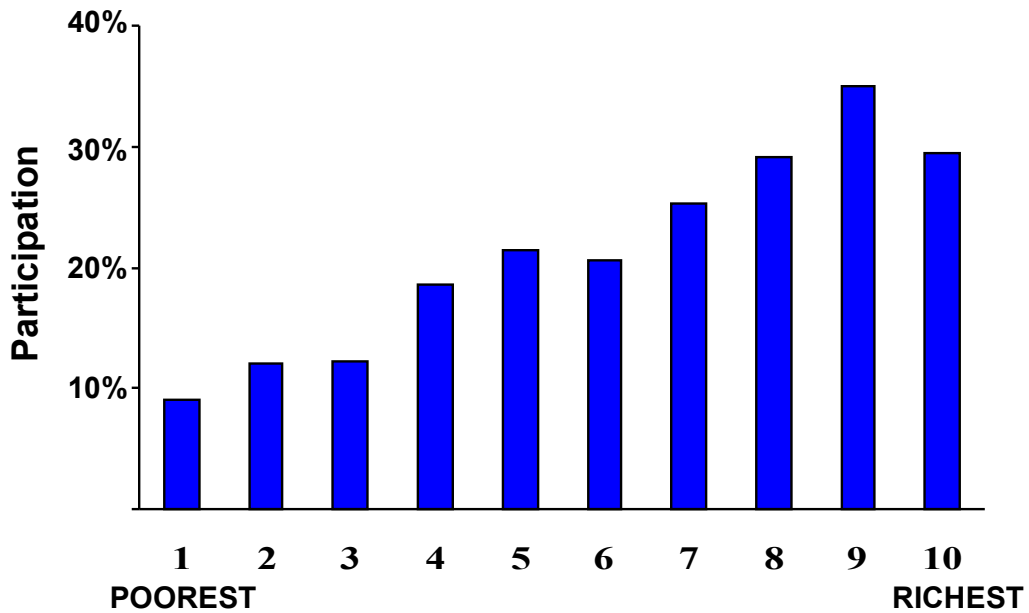
Nicol, A. R., Stretch, D. D, Fundudis, T., Smith, I., Davison, I. (1987). The nature of mother and toddler problems – I. Development of a multiple criterion screen. *Journal of Child Psychology and Psychiatry*, **28**, 739-754.

Wolke, D., Waylen, A., Samara, M., Steer, C., Goodman, R., Ford, T., Lamberts, K. (2009) Does selective dropout in longitudinal studies lead to biased prediction of behaviour disorders? *British Journal of Psychiatry*, **195**, 249 - 256.

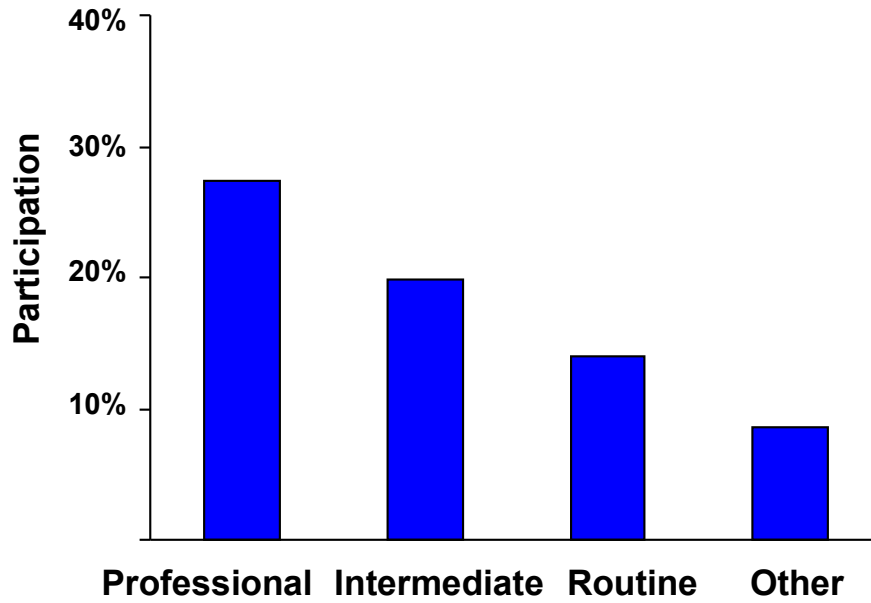
**Participation by Area Deprivation (Quintiles)**



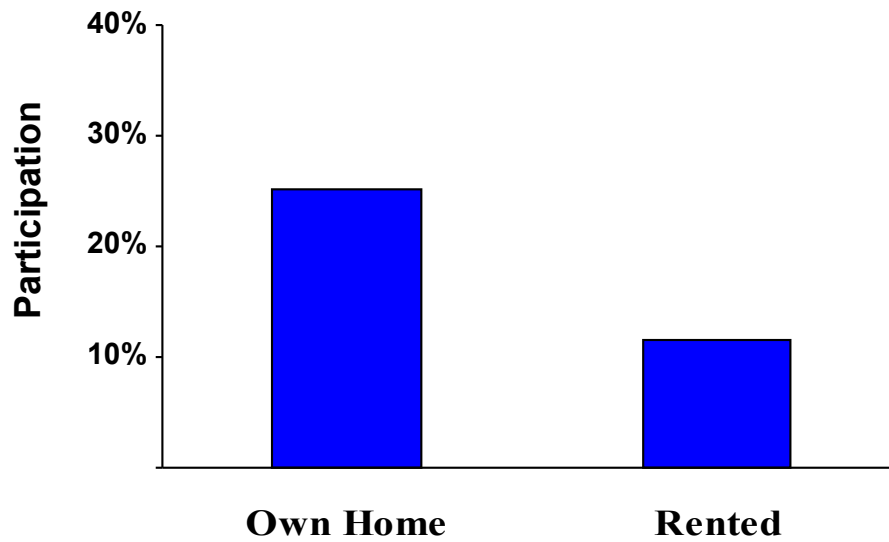
**Participation by Household Income (Deciles)**



**Participation by Social Class (NS-SEC 3)**

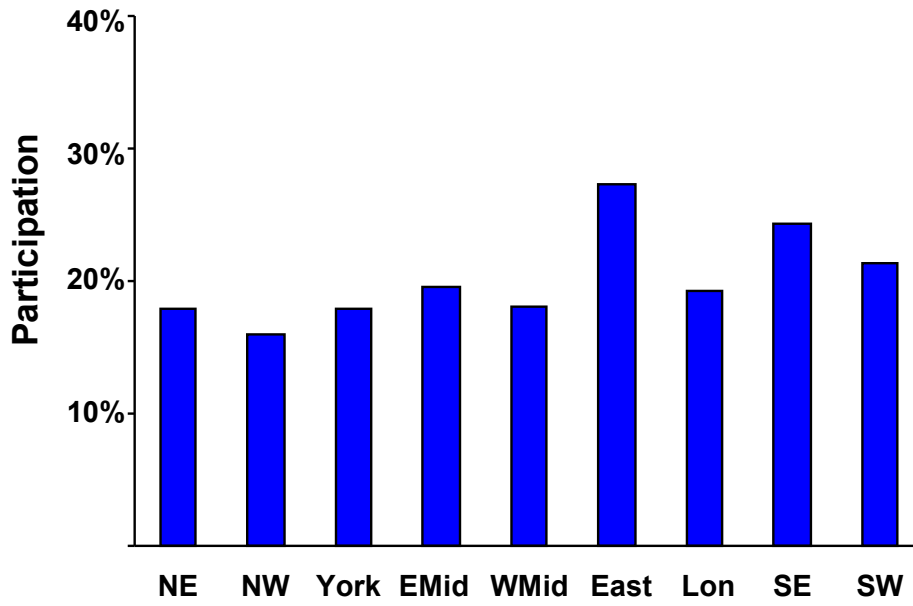


**Participation by Housing Tenure**

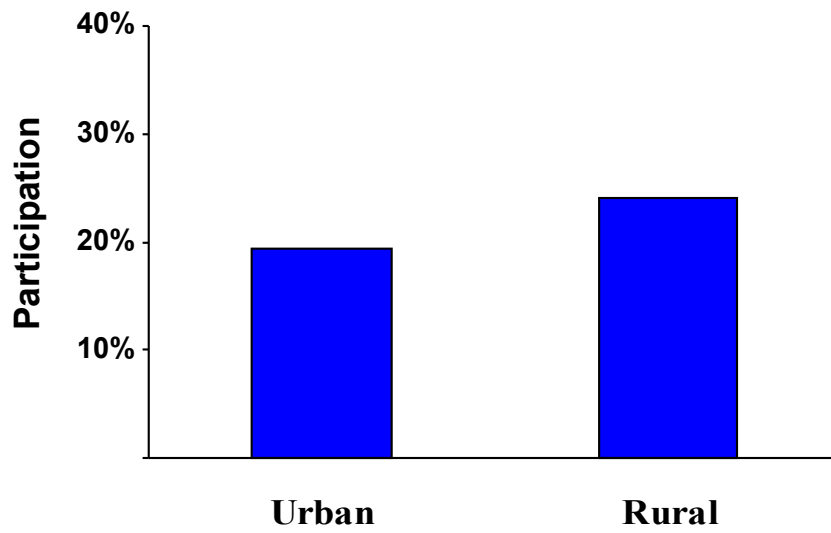




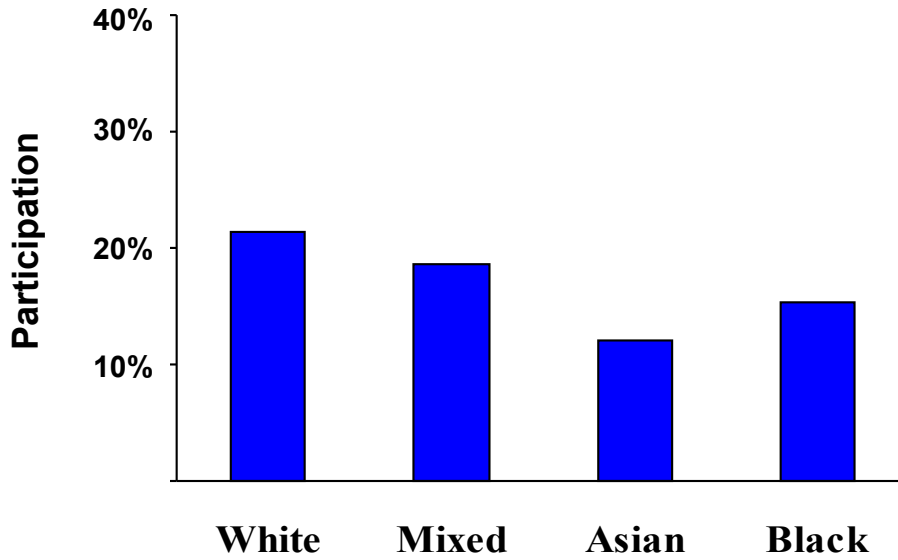
**Participation by region**



**Participation by urban or rural**



**Participation by ethnicity**



**Participation by family size**

