Project Protocol

(Frances Gardner, Stephen Scott, Sabine Landau, Andrew Pickles, Jennifer Beecham and Judy Hutchings).

1. Project title:

"How far could widespread dissemination of parenting programmes improve child antisocial behaviour and reduce social inequalities? Combining datasets from trials in different communities to establish for whom programmes are effective and cost effective"

2. Background: existing research and rationale for current study

There are compelling public health reasons for investing in parenting interventions. Poor parenting skills are strongly predictive of a wide range of poor mental and physical health outcomes in the next generation, especially youth antisocial behaviour (Ermisch, 2008; Hoeve et al, 2009); high quality parenting interventions are effective in the majority of cases and carry the potential to improve public health substantially. However, most trials have been of modest size and little systematic is known about their ability to bring about change in more disadvantaged groups. The proposed study will address this question by bringing to bear four important design features that increase power and value. Firstly, it will only analyse randomised controlled trials, thus overcoming the risks of bias that may arise from observational studies of public health strategies, where differences seen between populations do not necessarily translate into benefits when these are rigorously tested in intervention experiments. Secondly, rather than combining trials at the study level as is usual in meta-analytic studies using aggregate data, it will combine data from 14 trials at an individual participant level, thus greatly enhancing the opportunity to detect moderating (interaction) effects of social disadvantage. Our design represents a unique opportunity to overcome the problems of low power and reporting bias that beset subgroup analyses in most individual trials, and hence to enhance our understanding of how parenting interventions may reduce or widen health and social inequalities (Petticrew et al, 2011; Thompson & Higgins, 2005). Thirdly, it will draw upon qualitative research and public involvement, soliciting parents' views on factors affecting intervention success, to inform hypotheses for testing. Fourthly, it will apply cost and cost effectiveness approaches by health economists to enable potential benefits to public health to be predicted as accurately as possible.

Social inequalities are central here: both poor parenting and antisocial behaviour are strongly patterned by social disadvantage, and are linked to diminished life chances in key areas of schooling, employment, and health (Piquero et al, 2011; Waylen et al, 2008). UK cohort data show clear predictions from early poor parenting to child behaviour problems and diminished educational and health outcomes, suggesting it is a key mechanism for perpetuating health and social inequalities across generations (Ermisch, 2008; Kelly et al., 2011). Prospective studies (Belsky et al, 2009; Kim et al 2009) of high risk samples support the hypothesis that poor parenting mediates intergenerational transmission of adverse child outcomes. Furthermore, UK cohort analyses by applicants Pickles and Gardner suggest that social inequalities in both child and parent mental health appear to be widening over time (Langton et al., 2012; Schepman et al 2011).

Persistent antisocial behaviour[1] is a major public health issue, not least because it is the most common mental health problem in children. Oppositional/ conduct disorders affect 5% of the population (ONS, 2004). There are high health burdens into adulthood, including a 5-10 fold risk of alcoholism, drug abuse, criminality, domestic violence, STIs, unemployment and early death (Fergusson et al., 2005; Odgers et al., 2007; Piquero et al, 2011). As with poor parenting, there is a strong association with social disadvantage, with a 4-5-fold rate of antisocial behaviour in the most disadvantaged groups (ONS, 2004). The extra public cost of antisocial children is £225,000 by age 27, and cost savings appear to apply to children with mild and moderate problems, as well as to those

[1] Note that here we use the term antisocial behaviour synonymously with conduct problems, as they refer to the same phenomena; severe cases meet criteria for conduct disorder, with oppositional defiant disorder being a subtype. Since the term conduct problems/disorder isn't always widely understood beyond the relatively narrow confines of mental health, we also use the term antisocial behaviour.

with more severe antisocial behaviour (Sainsbury Centre for Mental Health, 2009), 10 times that of controls (Scott et al, 2001). Parenting interventions potentially form an important public health strategy for preventing antisocial behaviour and other poor outcomes in children, for a number of reasons. Firstly, because the public health and financial burden of child antisocial behaviour and its later consequences are very high, it provides an excellent opportunity for early preventive intervention. Serious, enduring antisocial behaviours in adulthood nearly all began in childhood – fewer than 10% of persistent cases begin after age 18 (Moffitt et al, 2003). Secondly, NICE and Cochrane reviews of RCTs (Dretzke et al, 2005; Furlong et al, 2012) show clear effects of parenting interventions for preventing child antisocial behaviour problems, and for enhancing parent and child mental health. Many policy bodies worldwide have recognised this (e.g WHO, 2010, UNODC, 2009, CDC, 2009). The National Research Council and Institute of Medicine (NRC/IOM, 2009) has issued a strong call for early preventive intervention trials to answer how mental health disorders can be prevented. These calls are echoed in UK policy. Thus, the Department of Health (DH) state the need to promote evidence based parenting programmes in several 2011 policy documents (e.g. *No health without mental health: A cross-government mental health outcomes strategy*; *Talking Therapies: a four-year plan of action)*. The NICE HTA report (Dretzke et al, 2005) recommends parenting interventions; their meta-analysis found an effect size of 0.6 SD on child problem behaviour, with good long-term effects, an extremely worthwhile effect in public health terms. On 27 February 2013 NICE will launch its Full Guideline for prevention and management of antisocial behaviour and conduct disorders, a further recognition of the public health importance of this problem. The draft guidance (Dec 2012) confirmed the previous HTA analyses, and recommends the use of high quality evidence-based parenting programmes to prevent the development of antisocial behaviour and conduct disorders. Like the economic modelling study by members of our own group (Bonin et al, 2011), the draft NICE Guidance shows good financial returns from investment in evidence-based programmes.

Specifically to increase the availability of evidence based parenting interventions, through the DH, the government launched the Children & Young People's version of Increasing Access to Psychological Therapies (IAPT) in 2011, and in February 2012 the then Health Minister, Rt Hon Paul Burstow, announced an extra £22m funding to extend this. Currently, half the funding goes towards teaching evidence-based parent training to a wide range of health, social care, private and voluntary staff, in community, non-NHS settings, so that it will have wide public health impact. The DH IAPT initiative builds on the government's 2007 establishment of the National Academy for Parenting Practitioners, where over 4000 practitioners were trained in evidence-based parenting programmes (Scott, 2010). Because of this wide dissemination, it is becoming pressing to determine the effectiveness of the intervention across a range of social groups; the question is not a theoretical one.

Moreover, if these programmes are to make a difference to the lives of the most disadvantaged families, then we have to be sure that these families are accessing and benefiting from these interventions. A well-recognised danger is that interventions may sometimes have greater benefits for more advantaged families (Lorenc et al, 2012), sometimes termed the 'inverse care effect', as was found in the early Sure Start evaluations (Rutter, 2006). If this were the case, then parenting programmes might have the potential for increasing, rather than decreasing social inequalities. There is therefore a need to determine the likely health gain, and cost-effectiveness, if there were widespread dissemination of high quality programmes, for the most disadvantaged groups.

Our proposal is to combine individual level data from 14 randomised trials of perhaps the highest quality parenting programme in the world, one that is being widely rolled out throughout the UK. This is the Incredible Years (IY, Webster-Stratton and Reid, 2010a), a 12-session group parenting programme that reaches the highest quality rating on the Commissioning Toolkit (Scott, 2010), and has been identified as effective in many systematic reviews (Dretzke et al, 2005; Furlong et al, 2012, Barlow et al, 2012) for preventing antisocial behaviour in children, and improving parenting quality and parental mental health. It has received Government funding for training in England, Wales and Scotland, and as a result of its widespread dissemination, 7 community-based RCTs have been completed (London, Plymouth, Oxfordshire, Wales, Birmingham). However, although each of these trials was adequately powered to detect main effects, with sample sizes between 76 and 215, they were not powered to conduct secondary analyses which would provide crucial information about differential (interaction) effects on families with different levels of social disadvantage, and from different ethnic groups. Our set of 14 trials comprises all the trials of IY in Europe (total N, 1825 families), including 8 UK trials (total N, 988), and 6 in other European systems, with one trial each in Ireland, Norway, Sweden and Portugal, and two in the Netherlands.

Analyses are needed to determine which subgroups respond better and which are less responsive to parenting programmes, given considerable evidence of heterogeneity of effects (NICE 2013; Reyno & McGrath, 2006). If, for example, there is less benefit for those living in disadvantaged conditions (indexed by equity factors such as occupation, education, low family income, ethnicity, gender, rural vs urban; Welch et al, 2012), then steps need to be taken to vary the intervention to be effective for such groups. Because of the overall effectiveness of parenting programmes, identifying such determinants of response offers the opportunity to make a considerable difference to those living in poverty and reduce social inequality for the next generation, and increase access to good preventive and treatment intervention.

Our unique pooled data set will create a combined sample that is powered to detect intervention effect moderation by individual-level variables. One might consider this approach an 'individual-level meta-analysis'. However, it will tell us a great deal more than conventional meta-analysis (Lambert et al, 2002; Riley et al 2010), which has been approached in two main ways. The first method, combining estimates of effects using conventional aggregate data meta-analysis, has the drawback of allowing assessment of effect moderation only by trial-level characteristics. The authoritative paper by Lambert et al (2002) comparing these usual meta-analytic methods, *vs*. pooling individual data concluded *"Meta-analysis of summary data may be adequate when estimating a single pooled treatment effect or investigating study level characteristics. However, when interest lies in investigating whether patient characteristics are related to treatment, individual patient data analysis will generally be necessary to discover any such relationships."* An aggregate/ summary data analysis approach is likely to miss important moderating effects. For example, the mean age of children in a range of trials may be similar, and the average effect sizes may also be similar, so that using trial-level comparisons, such as meta-regression or subgroup analysis, age would not be predictive of intervention effect. However, often within trials there is age variation, and by combining them at an individual case level, we will have requisite information to see whether there is effect modification by age, and by socio-economic variables we plan to investigate. A further problem that stems from the use of trial-level predictors, is that predictors are often confounded with one another (Lipsey, 2003), making it hard to interpret their meaning. An example of this can be seen in the subgroup analyses in Furlong et al's review (2012, p221-2), where trials conducted in research settings, or with more affluent parents, were also those more likely to be conducted by the developer - all factors that tend to produce larger effect sizes.

A second, less commonly used meta-analytic approach to investigating moderators is one that synthesises across trials the findings of moderator or predictor analyses from published trial data (e.g. Reyno & McGrath, 2006). This approach has the advantage of making use of within-trial variability in socio-economic characteristics, mitigating somewhat the problem termed as 'those confounded moderators' by Lipsey (2003), and does not require labour-intensive pooling of individual data. However, it suffers from serious drawbacks, leading to researchers recommending against its use (Brown 2011; Shadish & Sweeney, 1991, p889). A key drawback is that most trials do not report moderator data, raising the possibility of reporting bias, or at best, resulting in meta-analyses that can only summarise an incomplete picture. As well as the fact that trial outcome data is rarely broken down by equity factors, where trials do test socio-economic or other predictors, statistical models are specified in varying ways (Brown et al 2011), for example, some calculating interaction effects, but others only within-group predictors, rendering synthesis meaningless (Petticrew et al 2011; Welch et al 2012). These problems apply no less to parenting intervention trials (Furlong et al, 2012; Gardner et al, 2010), and can be overcome by use of pooled data.

Our strategy will largely overcome the drawbacks of these other approaches; pooling individual-level data is an exciting new approach to synthesis (Cooper & Patall, 2009; Brown et al, 2011), increasingly common in medicine in recent years (Riley et al, 2010), but rarely used in public health or psychosocial fields. Our approach is described in more detail in the methods and analysis sections.

2.1 Research questions

The proposed study would address the following three questions:
(1) to what extent do parenting interventions benefit the most socially disadvantaged families?
(2) what are the wider public health benefits and potential harms of parenting interventions?
(3) what mediates the effects on child outcome?

Question 1: To what extent do parenting interventions benefit the most socially disadvantaged families?

Our analyses would address many of the major indices of social disadvantage:

Question 1a: Family poverty and disadvantage

Despite a considerable number of high quality trials and systematic reviews on this topic, several vital questions about public health benefit of parenting interventions remain, that cannot be answered from individual trials, or systematic reviews, alone. First, it is unclear whether they are as effective at engaging and helping disadvantaged families, compared to more advantaged families. This is a crucial question for i) understanding impact on social inequalities, and ii) modifying intervention content and delivery methods to enable fair access. A number of trials and systematic reviews of parenting interventions have found weaker effects for more disadvantaged families (Lundahl et al, 2006; Reyno & McGrath, 2006). On the other hand, one review (Furlong et al, 2012) and some of the few individual trials testing moderator effects (Gardner et al, 2009; 2010; McGilloway et al, 2012) find that parenting interventions are just as effective, or more so, for the most disadvantaged families, who, without intervention, tend to do worse, suggesting the potential for reversing some of the poorer child outcomes associated with family poverty. Most trials have not used their data to ask these questions. Conflicting results in trials and reviews may be due in part to low power, and (in reviews) to the limitations of assessing these variables at trial level. In our pooled data set, all trials have several indices of family poverty and disadvantage. Further, the trials have data that will enable us to explore the mechanisms through which disadvantage may operate. This includes information on the number of sessions attended, which is one possible reason why families under stress may not benefit, since they may not access the intervention. The trials also have data on changes in parenting practices, which is an alternative mechanism through which disadvantage may operate, as the families may be living with multiple daily stresses that make it harder to enact new strategies. Thirdly, we have data on child characteristics associated with disadvantage (see below) which may reduce the effectiveness of programmes, despite adequate attendance and sufficient changes in parenting practices. We also know that contextual variables outside the family have a considerable influence on parenting styles and child behaviour, for example neighbourhood effects (Ingoldsby & Shaw 2002; Sampson & Morenoff, 2004). In our UK trials, we will explore moderating influence of neighbourhood deprivation and crime level on child outcome, using ward-level multiple deprivation indices (Noble et al, 2006), and police area data, linked to postcodes for each family.
Contextual effects that operate at the trial level, (eg, UK region, type of service provider) are covered in section 1d.

As well as drawing upon quantitative studies, we will test hypotheses that arise from qualitative studies of parents' experiences. A systematic review by Kane et al (2007), found 5 qualitative studies of the views of (mainly disadvantaged) parents who had participated in parenting interventions; our searches found several more recent studies of IY, two of which were embedded within a trial (Furlong & McGilloway, 2012; Morch et al, 2004). Barriers to uptake and success in intervention included stresses related to time pressure, financial pressure, the influence of antisocial neighbourhoods, reluctance to share problems with others, lack of support from family members. These findings show congruence between parents' views of barriers, and those factors drawn from the literature on moderators, and on risk factors for child antisocial behaviour (Murray & Farrington, 2010). We will

therefore examine these influences wherever they have been measured in the contributing trials, which is the case for neighbourhood characteristics, social support and income poverty. We will also be able to pool parent satisfaction data from 12 of our trials (see Question 2a, below).

Question 1b: Ethnicity

Few trials in the UK have been able to examine effects of parenting interventions on families from different ethnic backgrounds. This is vital in order to assess whether such services are likely to reduce or widen inequalities by ethnicity in child and maternal outcomes (Brooks-Gunn & Markham, 2006). It is becoming increasingly important in the UK, and for example recent population data in London suggest that in some of the more disadvantaged boroughs a third of children or more (over half in Tower Hamlets) belong to an ethnic minority. Where there is evidence from other countries, mainly from the US, the picture is quite mixed. Measuring a very wide range of parent and child outcomes, as well as parent engagement and satisfaction, Reid et al (2001) found surprisingly little evidence of differential effects of the Incredible Years parenting intervention by ethnicity; where there were differences by ethnicity, they tended towards greater engagement and uptake by some minority groups. On the other hand, much theoretical and prevention literature from the US focuses on the need for interventions to be specially adapted for different ethnic groups (Kumpfer et al. 2002; Castro et al, 2010). Even leaving aside the controversial question of which is most effective (Huey & Polo, 2008) such approaches would imply running parenting groups that are separated by ethnicity, and raises critical questions about what would be appropriate service delivery patterns for multi-ethnic UK inner cities (Moran et al, 2004). There have been very few studies of outcome differences by ethnicity in UK parenting trials, nor qualitative studies alongside trials. One exception is Scott et al's (2010b) trial conducted in a highly deprived London borough. They found considerable baseline differences in parenting practices by ethnicity, but intriguingly, no ethnic differences in attendance, or intervention effects on parenting skills. This trial therefore suggested that despite large initial differences, parenting programs based on 'Western' family values were equally effective with ethnic minority parents, when sensitively delivered, using a programme with an underlying philosophy that is collaborative and parent- centred (Webster Stratton, 2009). A Manchester study of parents' views of a similar programme, Triple P (Patel et al, 2011) suggested that Asian and African parents were most inclined to take up a parenting intervention, but white and African-Caribbean parents somewhat less so. The detailed ethnicity data from our London and Birmingham trials means that, when combined, we will be able to take a culturally more nuanced approach to examining ethnic differences in outcomes, uptake and parent satisfaction. There is extensive data available from all of these trials, using standard classifications of ethnicity. Parents self-classify using a large number of categories, which can be summarised as white British, black African, black African Caribbean, Indian subcontinent, East Asian, and Middle Eastern. It will be possible to code the data in terms of these broad categories, as well as by ethnic minority versus majority. The data potentially allow for more fine-grained classifications that these; however we will be cautious about deriving more than a small number of categories in order to ensure that was enough power for analysis.

With the exception of the two trials from Holland, in most of the studies from other European countries, numbers of ethnic minority parents are very small. In the two Dutch trials, approximately 70% of the parents are from ethnic minorities, including mainly immigrant families from North and East Africa, Turkey, Middle East and the Caribbean (around 190 minority families). Across the pooled data set, some 30% of the total sample would comprise families from ethnic minorities.

Question 1c: Child characteristics and needs.

Relatively little is known about how child characteristics, such as age, gender, and initial severity of behavioural problems, influence the effects of parenting interventions.

i. Age: Whether intervention effects (and cost effectiveness) vary by age of the child is a particularly salient policy issue. Current policy thrust is toward early parenting intervention, with the Allen (2011) Report on Early Intervention strikingly proposing that resources should be taken away from later intervention and redeployed to earlier age groups. Yet there is surprisingly little conclusive data on the

most effective age for targeting preventive interventions, with small trials providing conflicting results. For example, one recent trial found no age effects (McGilloway et al, 2012), another found a slight advantage of younger age (Gardner et al, 2010), but was limited by including only a narrow pre-school age range in the trial. Systematic reviews have also produced mixed findings; two reviews of age effects on parenting interventions found no differential advantage of young age (Lundahl et al, 2006; Furlong et al, 2012), and two found greater effects for older children (Serketich & Dumas, 1995; Weisz et al, 1995). However, most of these reviews are not up to date, and are severely constrained by lacking data on age at an individual (rather than trial) level. Moreover they are unable to control for baseline severity of child problems, at an individual level, vital because age is often confounded with severity (as it is with gender, older children and boys tending to have more severe problems), and there is evidence that children with more severe behaviour problems may gain more from these interventions. The trials that we propose to analyse have a wide age range, from 1-10 years.

ii. Gender: Gender effects on outcome will be examined. The picture from existing literature is complex: where girls present with severe antisocial behaviour problems, they often show more marked co-morbidity than boys. However, in prevention samples, they often have less severe behaviour problems, and this might contribute to finding stronger intervention effects in boys in some studies (e.g. Gardner et al's (2010) moderator analyses in Wales Sure Start Trial). However, other prevention trials find no such gender effects (Beauchaine et al., 2005; Conduct Problems Prevention Research Group, 2002). Therefore it is important that we analyse in a larger sample whether gender moderates intervention outcomes, while also controlling for initial severity and co-morbidity, in case these are associated with gender. If there are weaker effects for girls, programmes may need adjusting to take account of their needs.

iii. Initial severity of behaviour problems. Systematic reviews of parenting interventions again provide conflicting results, albeit based on contrasting meta-analytic methods for synthesising moderator effects. One review finds that children with higher levels of behaviour problems do better (Lundahl et al, 2006), another that they do worse (Reyno & McGrath, 2006), and a third (Furlong et al, 2012) finds no difference. With our pooled dataset, we will attempt to disentangle effectiveness of the intervention according to initial severity, and if there is an effect, test if this is accounted for by better attendance, which is often associated with greater severity, or greater parenting change. This will interact with cost effectiveness, since, due to the poor outcomes of the more severe group when untreated, taking them on is likely to be more cost-effective than taking on milder cases. The latter will have fewer cases that will go on to be very costly, and will probably change less.

iv. Comorbid child problems. Linked to initial severity, is the question of whether child co-morbid problems moderate intervention effects. Some studies suggest that children with high levels of other mental health problems (eg ADHD) do less well in parent training, but others have found as good a response for these children (Jones et al., 2008; Webster Stratton et al, 2011). If severe ADHD does moderate treatment response adversely, then this might suggest that before parent training is undertaken with this population, stimulant medication (as recommended by NICE) should be considered. We will also look at the impact of comorbid emotional problems, such as anxiety and depression, which for example reduce the effectiveness of some interventions for ADHD, so are worth examining in the context of antisocial behaviour/conduct problems.

Question 1d: Interactions between contextual variables and intervention effects:

There are many important contextual variables that can be coded at the level of the trial, that the literature suggests are likely to affect implementation and effectiveness of parenting interventions (Hutchings et al, 2007b; Scott, 2010). These include:

- type of service provider organisation (eg NGO vs statutory agency)
- profession delivering intervention/ level of professional training
- level of attention to fidelity of implementation
- university efficacy vs 'real-world' service effectiveness setting
- geographic factors: UK region; inner city vs small town, vs rural; UK vs other country.

Importantly, our pooled data will allow us to examine in the same models, how outcomes are moderated by both trial-level contextual, and individual-level variables (see Statistical Analysis Plan).

Thus for example, we will be able to explore whether any reduction in effectiveness is better accounted for by family level variables such as income and lone parenthood, or whether even after taking these into account, there are still service context, regional, rural vs urban location, or other trial-level effects.

Question 2: What are the wider public health benefits and potential harms of parenting interventions?

In addition to assessing child antisocial behaviour, the primary outcome, all of our trials assess a range of secondary outcomes that reflect the wider benefits of parenting interventions for family well-being. Firstly, these include improving parenting skill and parent-child relationships, with increases in positive involvement with children, and reductions in harsh parenting and abusive practices. Although these are termed secondary outcomes in most trials, they are also seen as crucial mediators between intervention and outcome (Gardner et al, 2010). Secondly, the programmes have been shown to improve adult mental health and well-being, including parental depression, confidence in their ability to be a successful parent, and improved partner relationships. Thirdly, some studies show generalisation to improved behaviour of other children in the family.

Harsh parenting is of particular importance for child well-being and quality of life; in prevention trials where many of the children show quite low levels of behaviour problems, these interventions impact public health by reducing levels of harsh or abusive parenting, and family stress. This has been found in universal and selective prevention trials (Prinz et al, 2009; Webster Stratton & Reid 2010a), and in studies of parents at high risk for abusive parenting (Barlow et al, 2006; Webster Stratton & Reid, 2010b). The Triple P trial (Prinz et al, 2009) showed that widespread implementation of a similar parenting programme reduced admissions to hospital for abuse, measured by county-level indicators. Recent reviews confirm that even mildly harsh parenting is associated with harmful biological effects on children, including for example dysfunctional cortisol secretion patterns, and raised C reactive protein which in turn is associated with increased cardiovascular disease and mortality (Scott, 2012).

Parent depression has been shown in some trials to be improved by parenting interventions, and this will be an important public health benefit to document. For parents with young children, many or most of their waking hours are spent caring for them, and qualitative studies suggest that failure to succeed in controlling child behaviour is a major source of lack of confidence and depressive cognitions (Morch et al 2004). Thus we will examine overall main effects, and possible impact in subgroups.

Question 2a. Service-user satisfaction: These important data on parents' views are rarely presented in detail in trial reports, or reviews (Furlong et al, 2012). Twelve trials have collected parent satisfaction data, but these have never been synthesised. Hence our study provides a unique opportunity to bring together these data, and examine differential perceived benefits by socioeconomic status and ethnicity.

Question 2b Harms: As well as benefits, it is important to consider potential harms, especially as they are rarely studied in parenting intervention trials. Where harms have been studied in psycho-social trials, they have mainly involved youth-focused interventions, and define harm as finding main effects in the unintended direction (Dishion et al, 1999; Petrosino et al, 2002). Recent systematic reviews of parenting interventions have not found evidence of harmful effects defined in this way (Furlong et al, 2012). Parents rarely report potentially harmful outcomes in qualitative studies; Morch et al (2004) mention none, despite interviewing many parents for whom intervention was not successful. A few parents in Furlong & McGilloway's (2012) study were concerned about increased conflict with partners related to trying new parenting techniques, and the lack of privacy in group interventions which discuss family problems. Furlong et al's (2012) Cochrane review planned to examine two potential adverse effects, namely burden to families in attending (eg childcare issues), and increased family conflict, but found no studies reporting these outcomes. Given this weak state of evidence on harms from parenting interventions, we propose to conduct cautious, exploratory (hypothesis generating, rather than hypothesis driven) analyses of the following question:

Is there evidence of main effects in the adverse direction? Our pooled sample may pick up effects on primary or secondary outcomes, which were not detected in individual trials.

Question 3: Can we identify mediators of child outcome?
Mediator analyses aim to test key theories of change in interventions, examining which potential mechanisms best predict change in the primary outcome (Bonell et al, 2013; Kraemer et al, 2002). They are important in helping to identify which are the essential ingredients in an intervention, especially important when interventions are rolled out into non-specialist services, and busy providers may be tempted to shorten or dilute a manualised programme. Potential mediators can be identified in two ways: firstly, from basic research on the theory underlying the intervention; such studies stress the importance of overt parenting behaviour (Gardner et al, 2007; 2010; Hoeve et al, 2009). Secondly, qualitative studies have assisted with the generation of hypotheses about possible mediators. Parents prioritise the sense of social support, confidence and shared problem solving gained from the group, more than the skills gained (Morch, 2004; Furlong & McGilloway, 2012).Mediators have been tested in only three of the current trials with mixed results: positive, but not harsh parenting, mediated in two UK trials Gardner, 2006; 2010) and harsh parenting in Norway (Fossum et al, 2008), findings echoed in other trials of parenting interventions (Dishion et al, 2008; Kling et al, 2010). Improvement in parent confidence did not mediate outcome in Gardner et al's (2006) trial. It is not known why no mediation effect was found for harsh parenting or parent confidence, despite the fact that robust intervention effects were found on these outcomes. It may be due to low power or that there was no effect of the intervention on the proposed mediator, or of the proposed mediator on child outcomes. The huge advantage of pooling individual data will be to ensure a more precise and unbiased estimate of mediator effects.

2.2 Summary of the rationale and benefits of the proposed study
To summarise, this unique study based on pooled individual data will be the largest of its type in the world, and will considerably advance our understanding of differential intervention effects, and cost-effectiveness of parenting interventions for families with differing levels of social disadvantage and child risk factors. By doing this, it will help determine whether such programmes are likely to reduce or widen social inequalities. This is an important public health question due to the damaging and expensive effects of antisocial behaviour, and is of direct relevance to the NHS which is investing heavily in programmes to prevent this. It will generate a more precise and generalisable estimate of the wider benefits of parenting programmes, and on likely mechanisms of change. But perhaps most importantly, if there are groups for whom programmes work less well, it will stimulate change in working practices to try to improve availability and effectiveness of these programmes to such groups.
Specific benefits include:
i) Enhanced power to examine effects by social disadvantage, wider benefits, cost-benefits and harms: Given the conflicting findings from many small trials and meta-analyses on differential effects by social group and child factors, there is a clear need to increase power and generalisability by combining across trials and settings. Such pooling of individual child and parent data would be a considerable advance on what can be gained from meta-analysis, which suffers from the considerable drawback of subgrouping families by taking the mean for the whole trial, thus failing to make any use of the wide within-trial variability. The large sample makes it more likely we will detect any potentially harmful effects, as well as wider benefits.
ii) Reducing reporting bias: Importantly, by pooling all available baseline and outcome variables across trials, our analyses will help to reduce selective reporting and publication bias, whereby positive secondary outcomes are reported more than those showing null or harmful effects, or where non-significant moderator analyses (if conducted) may potentially not be published (Brown et al, 2011). Reporting bias is known to be a considerable problem in many areas of health care; systematic

reviews find it to be linked to higher effect sizes (Dwan et al, 2008; Sterne et al, 2008). It is especially problematic in this field, where there are typically multiple secondary outcomes, and multiple measures of the same construct within and between trials (Furlong et al, 2012).

iii) Wider generalisability across community service contexts and regions: Because the trials were conducted in a range of service settings (NGOs, Sure Start services, Day Nurseries, primary schools), samples and regions, inferences will be generalisable. This will also allow us to examine contextual effects on outcome, and their interaction with individual level factors. In addition to trials conducted in community services, we include one trial conducted in the NHS. This trial is highly similar to the others, as the intervention is identical and the sample is highly similar in terms of problem severity and demographic background (e.g., this sample too includes large numbers of socioeconomically disadvantaged and ethnic minority families).

iv) Evidence is up to date: It is unlikely we have missed any European IY trials, as we have conducted extensive literature searches and expert contacting, which revealed many trials nearly and recently completed, and contacted IY, Seattle, who log all training for IY implementation. Of 14 trials, 6 are complete but not yet published, 2 were published in 2012; 5 in 2006-10.

3. Research objectives:

i) To combine data from 14 completed randomised trials (N=1798 families) of a community-based parenting intervention ('Incredible Years') aimed at preventing child antisocial behaviour, and improving poor parenting skill, problems that disproportionately affect families living in poverty. Family, maternal and child data will be pooled from all 14 trials, and cost data from 7 trials.

ii) To examine, using this unique pooled data set, whether the intervention is less or more effective and cost-effective at improving child and parental outcomes in the most disadvantaged families, compared to more average families, sub-grouping families by SES, poverty and lone parent status. We will also examine whether there are differential effects by family ethnicity, by rural vs urban location, and by child characteristics including age, gender, and initial severity of behavioural problems, and examine the extent to which these may be explained by trial level contextual factors.

iii) To examine wider benefits, and potential harms, of parenting interventions, including effects on harsh and positive parenting, parental depression, parenting confidence, partner relationships, child co-morbid problems (such as ADHD) and behaviour of other children in the family.

iv) To examine cost and cost effectiveness in the short term, and potential long-term cost-benefit generated by the intervention in the teen and adult years, by developing a model of long term impacts in the UK.

4. Research design (Please see p.16 for a list of trials; see uploaded trials chart for summary details of each trial).

Secondary analysis of 14 randomised controlled trials with 1798 families, pooling individual level data of child antisocial behaviour (primary outcome), parenting practices and parental mental health (secondary outcomes) and health economic measures. The potential for parenting programmes to reduce social disadvantage will be tested through moderator analyses that test differential effects measured by a range of important socio-economic indices, including income, parental education and lone parent status, and ethnicity, as well as child characteristics including age gender, severity of behaviour problems and comorbid problems such as ADHD. Our analyses will examine the extent to which these may be explained by trial level contextual factors.

4.1 Protection from risk of bias.

We have data to judge the risk of bias in the included trials: of the 14 completed trials, 5 were assessed for risk of bias in a recent Cochrane review (Furlong et al, 2012); 3 more were assessed by a systematic review of cross-country transportability of parenting evidence (Gardner et al., 2013).

Trials were assessed as mainly having low risk of bias, although in a few cases there was high or unclear risk in relation to attrition. Our pooled secondary analyses will be further able to reduce risk of bias, in the areas of bias due to attrition and - harder to detect in reviews - outcome reporting bias.

5.0 Study Population

Our sample consists of young children from the general population who are at risk of antisocial behaviour and their parents; most are from low income families, but with a good range of levels of social disadvantage. The 14 pooled trials have sample sizes ranging from 62-215, total N 1798. There are multiple trial sites in mainly low income areas, including 7 non-NHS sites in the UK, one NHS site in London, and 6 additional European sites, in Ireland, Norway, Sweden and Portugal, plus two in the Netherlands.

Inclusion and exclusion criteria: It is one of the strengths of the study that there is a range of contexts, and inclusion and exclusion criteria, yielding a range of child and family characteristics, both between and within trials. Thus, children's ages range from 1-10 years, with one trial focusing on toddlers (1-3 yrs), 3 on preschoolers (3-5 years), 3 on 4-6 year olds, and 7 on a wider age range, spanning preschool and middle childhood. Sample severity and parental help-seeking also vary; thus 7 of the trials conducted screening in services in low-income communities, followed by inviting parents to take part who rated their children as showing elevated levels of behaviour problems on a standardised instrument. One trial invited all families of toddlers in highly deprived areas; and 4 trials recruited families with more severe problems, who had been referred into NGOs and other community services, for help with conduct problems. One trial was conducted at the NHS for families referred for help with child conduct problems. One Dutch study recruited a very high risk sample, namely mothers of young children recently released from prison. Across the 14 trials, the proportion of lone parents ranged from 10-74%, and in the UK trials, the proportion of families on benefits ranged from 30-70%. There were 5 UK and 2 Dutch trials with a sizable proportion of ethnic minority families, ranging from 24-75%. Most of the trials operated very few exclusion criteria, other than those based on child age, or reaching a screening cut off. Where there were exclusion criteria, these were ones that would apply typically to usual service delivery, for example, several trials excluded children with severe learning disabilities. Thus the trials are representative of 'real life' public health practice. All trials had usual services as the comparator or group, rather than alternative programme or psycho-educational intervention. Therefore findings are likely to generalise to the addition of parenting programmes in settings similar to those of the trials.

6.0 Socio-economic variables

This study's main purpose is to investigate the extent to which parenting programmes can overcome social disadvantage, which will be measured using several important indices**.** Please see sections1a to 1d addressing Question 1 in the Background for details of these concepts. Here we note the data completeness:

Family socio-economic: All trials have data on percent of parents unemployed; lone parent; education level. Data on family welfare benefits is available for all UK, Irish trials, one Dutch trial; data on income, or SES based on parent job, is available in 9 trials.

Family ethnicity: All trials have data on percent of families from ethnic minorities; 7 trials (including the 4 largest) have significant ethnic minority participation (24 -76%, n= 530), and breakdown by multiple categories of ethnicity.

7.0 Planned interventions

The intervention is the Incredible Years (IY) basic parent programme, delivered by two group leaders to groups of 6-15 parents, in weekly session of 2 hours, for 12-14 weeks. All trials include data on attendance and loss to follow up.

8.0 Outcome measures

8.1 Primary outcome:

Child antisocial behaviour/conduct problems: measured by standardised parent self-report instrument. Most trials use one of two well-validated questionnaires, Eyberg Child Behavior Inventory (Robinson et al, 1980) or SDQ (Goodman, 2001). Eight trials also use a well-validated standardised semi-structured parent interview (Parent Account of Child Symptoms (PACS, Taylor et al, 1986), or Kiddie SADS (Kaufman et al 1997). Eight trials have observational measures of child behaviour problems.

8.2 Secondary outcomes measuring wider benefits of parenting programmes:

1. *Parenting behaviour – positive and harsh*: all trials; by standardised parent questionnaire in 10 trials (eg O'Leary Parenting Scale; Arnold et al, 1993); by direct observation in 8 trials.

2. *Parent mental health:* Twelve trials use well-validated instruments: Beck Depression Inventory (Beck et al, 1961; 5 trials), GHQ (Goldberg et al, 1997; 3 trials), PSI (Parenting Stress Index, Abidin,1990, 1 trial), SCL (Derogatis et al, 1973, 3 trials) WEMWBS (Tennant et al, 2007; 1 trial)
3. *Child ADHD and emotional symptoms*: by standardised parent questionnaire (SDQ or Conners (1994) scale), in 12 trials.
4. *Other outcomes.* In addition, there are standardised measures of other variables in a few trials, that when combined will provide further useful estimates of wider benefit: sibling behaviour problems; social support; partner relationship, sense of confidence in parenting.

8.3 Data synthesis, data quality, and missing data

i) Data Synthesis. Scott, Pickles and Landau have carried out extensive piloting of methods for combining trials, based on 3 of Co-I Scott's trials. This has allowed them to become familiar with challenges in combining data sets, and to develop strategies for moderation and mediation assessments and to ensure that we make realistic estimates of staff time needed for these considerable tasks. Our time estimates are informed also by the work of Beecham et al (2001), who analysed pooled mental health data. For the primary outcome data, using standardised questionnaires, pooling is relatively simple and Z scores against published norms will be used to make the Eyberg and SDQ questionnaires directly comparable, as is standard in more conventional meta-analysis. However, rendering interview measures such as socio-demographic variables directly comparable for individual trial data combination is less straightforward, as different scales and definitions have been used. In our pilot work we have shown that this can be done, but is surprisingly time-consuming. Likewise direct observational measures of positive and harsh parenting have used two main coding schemes (BCS, Aspland & Gardner 2003, DPICS, Eyberg & Robinson, 1981) where decision rules on combination of subscales need to be carefully checked for comparability.

ii) Data quality. There are three main sources of data. The questionnaires are standardised and so provided that parents have filled them in conscientiously, quality is good. The interviews were semi-structured, with extensive training of researchers and good inter-rater reliabilities (intra-class correlations 0.78 to 0.9 for PACS), thus providing good quality data (Le Couteur & Gardner, 2008). The direct observations used state-of-the-art methodology and are generally considered to be the 'gold standard' of measurement, often lacking in many psychosocial intervention trials but used in a majority of these trials, thus adding considerably to the validity of our findings, as it removes potential reporter bias, due to having parents provide the main outcome data. Instead, using these methods, independent observers, who are blind to group membership, and trained to high reliability criteria, code the data on parent and child behaviour (Aspland & Gardner 2003; Gardner, 2000).

iii) Missing data. At time 1, pre-randomisation, all trials have more than 98% complete data. Post intervention, the mean missing data rate is about 12%. This will be accounted for using standard statistical methods, see analysis section below. As techniques for dealing with attrition and missing data vary in the assumptions that they make regarding the process that generates the missing data, we will carry out analyses that are feasible while making relatively few assumptions, and then explore the impact of departures from these assumptions using sensitivity analysis (see section 11).

8.4 Longer-term outcomes for modelling cost effectiveness.

Here we will model a range of assumptions on longer term cost effectiveness. We will make assumptions based on three models, firstly that the life course trajectory after treatment persists into adulthood (best case); secondly that the improve life course trajectory persist for 50% of cases (middle case); thirdly that all benefits will be lost after five years (worst-case). These will be based on several high quality longitudinal epidemiological studies that have formed the basis for several costing projections (Bonin et al, 2011). What will be new here will be estimating the differential effect according to social disadvantage. Thus for example in the event that parenting programmes proved to be effective in reducing the natural history of worse prognosis in disadvantaged populations, there would be greater cost savings in this group than in the more advantaged group. Economic estimates will be based on standardised costings for projected rates of alcoholism, drug use, criminality, partner

violence, STIs, teen pregnancy, unemployment, early death, using methods already developed by the team (see section 11).

9 Assessment and follow up:
9.1 All outcome measures above have been collected at baseline, and 6 months later, post intervention. Most trials have a further follow up at 12 or 18 months after baseline, and two after 5-6 years (London, Norway). Most of the trials have a waiting list control design, meaning that although we will examine maintenance of change over this time period, there is no longer a randomised comparison group at 12 or 18 months; In addition, as noted above under section 8d, we will model long term effects into adulthood.

9.2. Assessment of harms
As described above (see further details under 'Harms' in Question 2b, Section 2, Background), there is no tradition of reporting harms, and no standardised parameters for measuring harm for parenting interventions in the literature. In other youth intervention trials, harms tend to have been defined as detecting main effects in a non-beneficial direction (Dishion,1999). We propose to check for evidence of main effects in the direction of harm on all primary and secondary outcomes. Given the lack of literature to guide hypotheses about harm, our analyses will be cautious and primarily aimed at hypothesis generation.

10. Proposed sample size
Power calculations for the total sample size give >97% power for the interaction term when compared with the treatment and covariate main-effects-only model for a treatment arm difference in the covariate effect size on outcome of .15 SD (significance level 0.05).

11. Statistical analysis plan
All analyses will firstly be done using the total sample from 14 trials, and then repeated using the eight British trials to check the sensitivity of results to including non-UK European studies.

11.1 Moderator assessments. In order to assess intervention effect modification (moderation), outcomes of the combined sample (*n*=1798) will be modelled. Moderation modelling will be carried out for each of the primary and secondary outcomes (see above) separately. We will consider both putative moderators measured at the individual child or parent level (see list above), and at the trial level. As mentioned before the advantage of this individual level analysis over conventional aggregate data meta-regression is that it enables the assessment of intervention effect moderation by both trial-level and individual-level variables (Brown et al, 2011).

Random effects modelling assuming normally distributed outcomes will be used to separate individual-level variation from trial-level variation. Specifically for a given outcome, say child antisocial behaviour/conduct problems, the dependent variable will be post treatment child outcome (at the available post intervention time points). Trial-level random intercepts will represent trial heterogeneity in outcome; trial-level random coefficients of the intervention covariates will allow for intervention effect heterogeneity across trials (due to factors such as implementation differences or differences in trial target populations or general service organisation contexts affecting control groups); group-level intercepts will model training/group constitution effects of within the IY arm of each trial and child-level random intercepts will account for extra correlation due to two post randomisation outcome measures taken per child; and fixed explanatory variables will be:-

- pre-randomisation values of the outcome (eg baseline child conduct),
- covariate coding the follow-up time points (6; 12 or 18 months; and longer where present).
- covariate coding the randomisation group effect at the follow-up time points.
- the putative moderator under investigation
- interaction moderator x covariate coding the randomisation group effect at the follow-up time points

As tends to be standard practice in psychosocial RCT analyses, pre-randomisation values of the outcome variable are included in the model to gain precision for the intervention effect estimate. Here the parameters of main interest are those describing the change in the post intervention effect at 6, and 12 or 18 months, in response to varying the level of the putative moderator variable. We will test the statistical significance of this interaction and if detected we will describe its nature by estimating intervention effects within subgroups defined by the moderator. We will evaluate the amounts of between- and within-trial variability that can be explained by the moderation effect. Additional tests will be undertaken to check that significant moderation effects do not occur simply as the result of differential treatment compliance. We will also exploit the data to assess any moderation effect heterogeneity across trials (by letting the coefficient of the respective interaction term vary between trials). Most trials will not contribute 12-18 month data since they employed a waiting-list control design; the waiting list group received the intervention after the 6 month assessment and was no longer followed up. However, we will include all available data in the analysis, and if any intervention effect moderation is deemed not to vary between trials, then data from the IY arm in these trials can inform the assessment of intervention effect moderation at the 12 or 18 months time points.

The random effects models will be fitted using maximum likelihood and resulting inferences are valid in the presence of missing data provided that the process that generates missing values is missing at random (MAR). In the current context this means that only variables measured and included in the model, either as dependent or explanatory variables, can predict missingness. The assumption is less restrictive than the missing completely at random (MCAR) assumption made by methods traditionally employed in psychology (e.g. repeated measures ANOVA). In addition, the random effects modelling will be carried out for “long-format data” and thus will enable us to use all the available outcome measures. However, we are likely to encounter missing values in some of the putative baseline moderators under investigation, and moderator missingness may be predicted by post intervention child outcomes. To ensure that all cases with a post treatment outcome can contribute to the analysis and that analyses remain valid under this particular MAR process we use multiple imputation to impute missing values in pre randomisation variables. Finally, the impact of possible departures from MAR on our findings will be evaluated by means of sensitivity analysis (Carpenter et al, 2007; White et al, 2011).

11.2. Profile of wider public health benefits

The set of outcomes measuring wider public health aspects (see list of outcomes) will be modelled simultaneously for the combined sample (multivariate model). The long-format dependent variable will consist of the set of public health measures obtained post-intervention. Trial-varying random intercepts will again be used to account for trial heterogeneity in outcomes; trial-varying random coefficients of the intervention covariates will allow for intervention effect heterogeneity across trials; group-level intercepts will model training/group constitution effects of within the IY arm; subject-varying-intercepts will account for correlation between repeated measures; and an unstructured covariance matrix will be used to describe the co-variances between the different outcome measures; and fixed explanatory variable will be:

- pre-randomisation values of the respective outcome variables (captured by 6 variables each representing baseline values of the specific outcome variable and containing “0”s otherwise)
- outcome variable (5 dummy variables, references the outcome variable)
- covariate coding the follow-up time points (6; and 12 or 18 months)
- covariate coding the randomisation group effect at the follow-up time points for outcome variables 1-6

The objective of the analysis then is the estimation of the (multivariate) intervention effect on the outcome set. The results will be illustrated by means of a profile plot that displays estimated overall intervention effects and their confidence intervals against outcome variables and time. As for 11.1 some trials will not contribute 12 or 18 month data since they employed a waiting-list control design. We will again include all available data in the analysis, and if any intervention effect does not vary between trials then data from the IY arm can increase the precision of the (overall) intervention effect estimates at the follow-up time points. Models will assume multivariate normality for continuous outcome

variables or latent variables underlying binary outcomes (threshold model). Alternative models for multivariate repeated measures data will also be considered.

Models will again be fitted using maximum likelihood and the multivariate modelling will provide valid inferences in the presence of missing outcome values provided the missing value generating process is Missing at Random (MAR). Due to our analysis model simultaneously modelling all the outcome measures this assumption here allows observations on one outcome measure to predict missingness of another. The impact of possible departures from MAR on our findings will again be evaluated by means of sensitivity analysis (Carpenter et al, 2007; White et al, 2011).

Unlike in traditional approaches employed in psychology (MANOVA), families will contribute to our multivariate analysis provided at least one of the outcome variables is available at at least one of the post treatment time points. This reduces the risk of selection bias.

11.3. Mediation assessment

Where relevant parenting variables are available mediation analyses will be undertaken to determine which aspects of parenting are the essential mechanisms of change in child outcomes. Mediation analyses will aim to partition overall or subgroup-wise intervention effects into mediated and non-mediated components. We will use the Baron and Kenny (1986) approach to explore mediation.

Statistical analyses will be carried out in Stata using the user-contributed gllamm command (Rabe-Hesketh, Skrondal, & Pickles, 2005) and MPlus if necessary.

12. Economic analysis

Data preparation: Economic analysis on the combined individual-level dataset will reduce the uncertainty around cost-effectiveness findings from the single studies, which are commonly powered to the outcome analysis rather than costs. Central to the estimation of costs and cost-effectiveness are the records of service use for each person in the trial, including use made of the Incredible Years programme by the intervention groups. These data were recorded for six of the thirteen individual trials included in this proposal in the early-mid 2000s. The remaining trials were excluded as they were undertaken in non-UK countries for which the service array and public sector financing systems are very different, or were UK trials that did not collect any service use information. The service use records for each person is similar, and five studies used a variant of the *Client Service Receipt Inventory* (Beecham & Knapp, 1992) to record information on the frequency with which (mental) health, primary care and social care supports were used. Three of the studies include data on education supports, four contain information on the impact of the child's behaviour on parental employment, three have some data on parental use of services, and two have additional data on the impact on parental time spent caring for the child.

All service use data will be merged and cleaned by the economics team, ensuring that categories of service use and measures (type, frequency, duration) are comparable across all studies. Service utilisation rates for the baseline and first follow-up (6 months) will be described for each trial individually and for the combined sample, identifying any obvious differences between the trials. Our final decision about including the trial based in Ireland (where the public sector service array, financial and organisational context is slightly different) in the economic analyses will be made at this stage and following discussions with the researchers.

For each service used by participants, an appropriate unit cost will be obtained from publically available sources (e.g. Curtis 2011) or calculated using an equivalent approach (Beecham, 2000). We will also use a consistent approach to missing time values (this is most likely to be duration of visits)

using the median across the full merged sample. The total cost of service use will be calculated for each participant, as well as cost sub-totals for education support, (mental) health services, primary care and social care. This re-estimation of unit costs and per-participant costs will improve the comparability of costs across the trials by ensuring that any cost differences are not due to the individual approaches to estimating unit costs or total costs without deviating from the original data collection. Where data are available parental time costs (c350 families) and days of lost employment (c.500 families) will be calculated.

We will identify how each individual study has estimated the costs of the IY intervention and look at the estimations in the latest NICE guideline. However, to ensure comparability these costs are also likely to be re-calculated. We will employ the commonly used Service Information Schedule to record resources needed to provide the intervention, including staff time, overheads, materials and travel costs for each phase of the intervention (design, preparation and delivery; see Bonin & Beecham 2012). The intervention cost may vary between trials, as different elements of the Incredibly Years programme are delivered, but our approach will again avoid measurement error due to different approaches to intervention costing.

Data analysis: The economic analysis will consist of three broad activities
Analysis of cost data: Starting with the moderators identified by the analysis of outcomes, cost variations at baseline and follow-up will be explored. First, costs will be described and cost differences between sub-groups will be tested using uni-variate regression models or t-tests. Then, multi-variate models will be fitted. All statistical models will take into account the likely skewed distribution of cost data using regression-type methods such as Generalized Linear Models (Nelder & Wedderburn, 1972) or bootstrapping (Efron & Tibshirani, 1993) as appropriate. This will include testing whether costs are clustered within trials once individual level characteristics or indicators of social disadvantage and other co-variates have been taken into account. (This might be due, for example, by systematic differences in the supply or pattern of services in different locations.) The results of this analysis will inform the cost-effectiveness analysis and the decision-analytic model described below.

Cost-effectiveness analysis: This will be conducted for the combined data, employing a net-benefit regression framework, where the net monetary benefit for each individual is calculated by multiplying various levels of willingness to pay (WTP) for an improvement in outcome by the change in outcome and subtracting costs (O'Brien & Briggs, 2002; Glick, et al., 2007). The difference in net monetary benefit between the intervention and control group for each value of WTP is then estimated using a regression model, controlling for baseline costs, confounders and mediators identified above (Hoch, Briggs et al, 2002) and accounting for the distribution of cost data and clustering as necessary. The probability that the experimental group experienced a higher net monetary benefit than the control group, i.e. the probability that the intervention would be considered cost-effective, will then be plotted against the corresponding value of WTP, resulting in a cost-effectiveness acceptability curve (CEAC; van Hout et al, 1994). These analyses will take a public sector perspective focussing on the children's support and use one parent and one child outcome measure. This analysis will be repeated for sub-groups to identify relative cost-effectiveness for the sub-groups identified in the outcome analysis; by age group, gender, ethnicity, severity, and indicators of social disadvantage

Economic modelling: The results of the analyses described above will be used to update and expand our existing model of the longer-term benefits of parenting programmes for children at risk of developing persistent conduct disorders up to age 30 (Bonin et al. 2011). These data will improve our estimate of service costs associated with conduct problems, allowing us to distinguish by age, sex, baseline severity of behaviour problems and other factors identified in the analysis. The effectiveness findings at six months and the subsequent follow-up (commonly only available for the intervention group) will also be entered into the model. Together these will facilitate a more accurate estimate of the longer-term benefits of the intervention in terms of reduced costs to the public sector (health care, social care, criminal justice system), the voluntary sector and to victims of crimes.

Power Calculation

The sample size for the main public sector costs analysis will be c400 in the intervention and control groups, with a slightly smaller sample used when considering the costs of educational support. Data on parental impact (service use, time costs & lost employment) will inform analysis from a societal perspective. The power calculation for the economic analysis uses the formula for estimating sample size for net-benefit analysis provided by Willan (2001) and is based on the cost-effectiveness analysis of the Incredible Years programme by Edwards et al. (2007). The mean Eyberg Problem score at follow-up was 117.17 (SD 35.99) in the intervention and 140.74 (SD 40.77) in the control group. Average costs were £2,881 in the intervention and £523 in the control group. Standard deviations for costs were not reported, so we assumed that they were four times mean costs, £11,524 in the intervention and £2,092 in the control group. The correlation of costs and outcomes was assumed to be 0.5, with alpha=0.05 and beta=0.08 but the required sample size is robust to changes in positive correlation (i.e. higher costs for higher benefit) from 0.0-1.0. Under these assumptions, the sample size per group required to estimate a difference between groups in net benefit at a WTP of zero and 80% power is n=194, a total n of 388.

13. Ethical arrangements

The proposed project has been approved by relevant Research Ethics Committee at Oxford University, and the approval letter submitted to PHR. All data will be anonymised, and a data sharing agreement will be drawn up, before any data is shared between sites. To ensure the highest ethical standards, the project employs a conservative definition of anonymity. This includes removing not just names and addresses, but all potentially identifying information; each data site will remove or pre-code any information that might identify participants, such as postcodes, birthdates, or unusual family characteristics or events. The data sharing agreement will be based on a successful model used within PSSRU at LSE (where Prof Jennifer Beecham is based) for sharing NHS and social care data collected by the NHS Information Centre for health and social care (NHS IC), and local authorities. We will ensure that this document follows the best practice guidelines for researchers laid out by the UK data archive May 2011 - guidelines also followed by the MRC and ESRC. The agreement addresses many of the key ethical issues associated with sharing anonymised data, and it requires that data usage is restricted to a named set of persons. In addition, each PI and team member for the current project is bound by the data protection and ethical codes in operation in their institution.

Appendix: list of trials and reference list

<u>List of Incredible Years Trials (n=14) for pooled data set:</u>

i) England and Wales:

1. Morpeth, L, Berry, V, Blower, S, Tobin, Taylor, R, Edwards, RT, Linck, P, Bywater, T, Axford, N, & Lehtonen, M (2013). The effectiveness of the Incredible Years (IY) pre-school parenting programme in Birmingham, UK: A randomised controlled trial. (Investigators based at Dartington Social Research Centre).
2. Gardner F, Burton J, Klimes I. (2006) Randomised controlled trial of a parenting intervention in the voluntary sector for reducing conduct problems in children: outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry* 47:1123-32. (Site, Oxford)
3. Hutchings J, Bywater T, Daley D, Gardner F, et al. (2007). Parenting intervention in Sure Start for children at risk of developing conduct disorder: randomised controlled trial. *BMJ*:334-678.
4. Hutchings, J, Griffith N, Williams M, Gridley N. (2012). Evaluating the Incredible Years Toddler programme in disadvantaged Flying Start areas of Wales: A report to the Welsh Government. Bangor University, Centre for Evidence Based Early Intervention.
5. Scott S, O'Connor T, Futh A, Price J, Matias C & Doolan M. (2010). Impact of a parenting program in a high-risk, multi-ethnic community: The PALS trial. *Journal of Child Psychology and Psychiatry* 51, 1331-134.

6. Scott S, Sylva K, Doolan M, Price J, Jacobs B, Crook C, et al. (2010). Randomised controlled trial of parent groups for child antisocial behaviour targeting multiple risk factors: the SPOKES project. *Journal of Child Psychology & Psychiatry* 51:48-57
7. Scott S, Sylva K, Beckett C, Doolan M, Kallitsoglou A, Beecham J & Ford, T with the HCA study teams (2013, in press). Which type of parenting programme best improves child behaviour and reading? The Helping Children Achieve trial. Final Report to DCSF.
8. Scott S, Spender Q, Doolan, M, Jacobs, B, & Aspland, H (2001). Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *British Medical Journal, 323*, 194-200.

ii) Ireland and other European countries:

9. McGilloway, S., Ni Mhaille, G., Bywater, T., Leckey, Y., Kelly, P., Furlong, M., Comiskey, C., & Donnelly, M. (2012). A parenting intervention for childhood behavioral problems: A randomised controlled trial in disadvantaged community-based settings. *Journal of Consulting and Clinical Psychology, 80*, 116-27 (Ireland)
10. Axberg, U., & Broberg, A. G. (2012). Evaluation of the Incredible Years in Sweden: The transferability of an American parent-training program to Sweden. *Scandinavian Journal of Psychology,* 53, 224–232.
11. Larsson, B., &. Fossum, S, et al. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children: results of a randomized controlled trial. *European Child & Adolescent Psychiatry*,18, 42-52
12. Gaspar, M. &. Seabra-Santos, M. Randomised trial of the Incredible Years Basic Parenting Programme in Portugal (to be completed Spring 2013)
13. Leijten, P., Raaijmakers, M. A. J., Orobio de Castro, B., Van den Ban, E., & Matthys, W. (2013). Effectiveness of the Incredible Years parent training for the reduction of disruptive behavior in migrant children. (In preparation). (Netherlands)
14. Menting, A. T. A., Orobio de Castro, B., Matthys, W., & Wijngaards-de Meij, L. D. N. V. (2013). A randomized trial of parent training for mothers being released from incarceration. (Submitted for publication). (Netherlands)

References

Abidin RR (1990). *Parenting Stress Index*. Odessa: PAR.

Arnold DS, O'Leary SG, Wolff LS, Acker MM. (1993). The parenting scale: a measure of dysfunctional parenting in discipline situations. *Psychological Assessment,* 5, 137-44.

Allen G. (2011). *Early Intervention: The Next Steps*. Independent Report to Her Majesty's Government. London: HM Government.

Aspland, H. & Gardner, F. (2003). Observational measures of parent child interaction. *Child and Adolescent Mental Health, 8,* 136-144.

Axberg, U., & Broberg, A. G. (2012). Evaluation of The Incredible Years in Sweden: The Transferability of an American Parent-Training Program to Sweden. *Scandinavian Journal of Psychology 53,* 224–232.

Barlow J, Smailagic N, Huband N, Rollof V, Bennet C (2012). Group-based parenting programmes for improving parental psychosocial health, *Cochrane Database of Systematic Reviews* (6).

Barlow J, Johnston I, Kendrick D, Polnay L, Stewart-Brown S (2006). Systematic review of the effectiveness of parenting programmes in treating abusive parenting, *Cochrane Database of Systematic Reviews,* (3), 1 – 20.

Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Barrett, H. (2010). *The delivery of parenting skills training programmes*. London, Family & Parenting Institute.

Baydar N, Reid, M, Webster-Stratton, C. (2003). The role of mental health factors and program engagement in the effectiveness of a parenting program for Head Start mothers. *Child Development, 74*, 1433-1453.

Beauchaine, T., Webster-Stratton, C., & Reid, M. J. (2005). Mediators, moderators, and predictors of one-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis *Journal of Consulting and Clinical Psychology, 73*, 371-388

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-71.

Beecham, J. (2000). *Unit Costs - Not Exactly Child's Play. A Guide to Estimating Unit Costs for Children's Social Care*. DoH, Dartington Social Research Unit & PSSRU.

Beecham, J. & M. Knapp (1992). Costing psychiatric interventions. *Measuring Mental Health Needs.* G. Thornicroft. London, Royal College of Psychiatrists. 2nd edition 2000.

Beecham J, Astin J, Mummery K (2001). Linking datasets, people and resources. PSSRU Discussion Paper 1707, London: PSSRU, LSE.

Belsky, J., Conger, R. D., & Capaldi, D. M. (2009). The intergenerational transmission of parenting: Introduction to the special section. *Developmental Psychology, 45*, 1201-1204.

Bonell, C, Fletcher, A, Morton, M, Lorenc, T. (2013). Realist Randomised Controlled Trials: a new approach to evaluating complex public health interventions, *Social Science & Medicine*, 75, 299-306.

Bonin E & Beecham J (2012). Costing multi-site group-based CBT workshops, in L Curtis (2012) *Unit Costs of Health and Social Care 2012,* PSSRU, University of Kent

Bonin, E., Stevens, M et al. (2011). Costs and longer-term savings of parenting programmes for the prevention of persistent conduct disorder: a modelling study. *BMC Public Health* 11(803).

Brooks-Gunn, J & Markman, L. (2005). The Contribution of Parenting to Ethnic and Racial Gaps in School Readiness*. The Future of Children, 15,* (1) Special Issue*:* School Readiness: Closing Racial and Ethnic Gaps.

Brown CH, Sloboda Z, Faggiano F, Teasdale B, Keller F, Burkhart G et al. (2011). Methods for synthesizing findings on moderation effects across multiple randomized trials. *Prevention Science*, published online DOI 10.1007/s11121-011-0207-8.

Carpenter JR, Kenward MG, White IR. (2007). Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Stat Methods Med Res*;16, 259-75.

Castro, F., Barrera, M. & Steiker, L (2010). Issues and Challenges in the Design of Culturally Adapted Evidence-Based Interventions The cultural adaptation of prevention interventions: resolving tensions between fidelity and fit. *Annual Review of Clinical Psychology*, 6, 213–39

CDC (2009). *Parent Training Programs: Insight for Practitioners.* Atlanta, GA: Centers for Disease Control.

Conners CK. (1994). The Conners rating scales: use in clinical assessment, treatment planning and research. In: Maruish M, ed. *Use of psychological testing for treatment planning and outcome assessment.* Hillsdale, New Jersey: Erlbaum.

Cooper, H., & Patall, E. A. (2009). The relative benefits of metaanalysis conducted with individual participant data versus aggregated data*. Psychological Methods*, 14, 165–176.

Conduct Problems Prevention Research Group (2002). Predictor variables associated with positive Fast Track outcomes at end of third grade. *Journal of Abnormal Child Psychology*, 30, 19–36.

Curtis, L. (2011). *Unit Costs of Health and Social Care 2011*. Canterbury, Personal Social Services Research Unit, University of Kent.

Derogatis, L. R., Lipman, R. S. & Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale – preliminary report. *Psychopharmacology Bulletin*, 2, 17–36.

Dishion, T., McCord, J. & Poulin, F. (1999) When Interventions Harm: Peer Groups and Problem Behavior. *American Psychologist*, 54(9), 1-10.

Dishion, T. J., Shaw, D. S., Connell, A., Gardner, F., & Weaver, C. Wilson, M., (2008). The Family Check-Up with high-risk indigent families: preventing problem behavior by increasing parents' positive behavior support in early childhood. *Child Development, 79*, 1395-1414

Dretzke J, Frew E, Davenport C, Barlow J, et al. (2005). The effectiveness and cost-effectiveness of parent training/education programmes for the treatment of conduct disorder, including oppositional defiant disorder in children. *Health Technology Assessment* 9(50):III.

Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, et al. (2008) Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. *PLoS ONE 3*(8): e3081.

Edwards R. T., Ceilleachair A, Hutchings J, et al. (2007). Parenting programme for parents of children at risk of developing conduct disorder: cost effectiveness analysis. *BMJ* 334(682).

Efron, B. & R. Tibshirani (1993). *An Introduction to the Bootstrap*. NY, Chapman & Hall.

Emsley, R , Dunn G & White, I. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions *Statistical Methods in Medical Research*; 19: 237–270

Ermisch, J. (2008) Origins of social immobility and inequality: parenting and early child development. *National Institute Economic Review*; 205: 62-71

Eyberg SM, Robinson EA. (1981). Dyadic parent-child interaction coding system. Seattle, WA: Parenting Clinic, University of Washington.

Fergusson D, Horwood J, Ridder E. (2005). Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology & Psychiatry,* 46, 837-49.

Furlong M, McGilloway S, Bywater T, Hutchings J, Donnelly M, Smith S, O'Neill C (2012). Behavioural/cognitive-behavioural group-based parenting interventions for children age 3-12 with early onset conduct problems. *Cochrane Database of Systematic Reviews*.

Furlong, M., & McGilloway, S. (2012). The Incredible Years Parenting Program in Ireland: A qualitative analysis of the experience of parents living in disadvantaged areas. *Clinical Child Psychology and Psychiatry, 17*, 616-630.

Gardner F, Burton J, Klimes I. (2006). Randomised controlled trial of a parenting intervention in the voluntary sector for reducing conduct problems in children: outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry* 47:1123-32.

Gardner F, Hutchings J, Bywater, T, Whitaker, C. (2010). Who benefits and how does it work? Moderators and mediators of outcomes in a randomised trial of parenting interventions in multiple 'Sure Start' services. *Journal of Clinical Child & Adolescent Psychology*, *39, 568-80.*

Gardner, F., Connell, A., Trentacosta, C, Shaw, D.S., Dishion, T. J., Wilson, M. (2009). Moderators of outcome in a brief family-centred intervention for preventing early problem behaviour. *Journal of Consulting and Clinical Psychology, 77,* 543-553.

Gardner, F. Shaw, D., Dishion, T., Burton, J. & Supplee, L. (2007). Randomised trial of a family-centred approach to preventing conduct problems: Effects on proactive parenting and links to toddler disruptive behaviour*. Journal of Family Psychology, 21, 3*98-406.

Gardner, F. (2000). Methodological issues in the direct observation of parent-child interaction: do observational findings reflect the natural behaviour of participants? *Clinical Child & Family Psychology Review, 3*, 185-198.

Gardner F, Montgomery P & Knerr W. (2013). Transportability of parenting programmes across cultures and countries: A systematic review. In preparation.

Glick, H., J. Doshi, et al. (2007). *Economic Evaluation in Clinical Trials*. Oxford, OUP.

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

Goldberg, D.P., Gater, R., Sartorius, N., Ustun, T.B., Piccinelli, M., Gureje, O., et al. (1997). The validity of two versions of the GHQ. *Psychological Medicine, 27*, 191–197.

Hoch, J. S., A. H. Briggs, et al. (2002). Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ* 11, 415-430.

Hoeve M, Dubas J, Eichelsheim V, et al. (2009). The relationship between parenting and delinquency: A meta-analysis. *Journal of Abnormal Child Psychology* 37:749-75.

Huey SJ, & Polo AJ. (2008). Evidence-based psychosocial treatments for ethnic minority youth. *Journal of Clinical Child & Adolescent Psychology* 37, 262–301

Hutchings J, Bywater T, Daley D, Gardner F, et al. (2007a). Parenting intervention in Sure Start for children at risk of developing conduct disorder: randomised controlled trial. *BMJ*:334-678.

Hutchings, J., Bywater, T., & Daley, D. (2007b). Early prevention of conduct disorder: how and why did the North and Mid Wales Sure Start study work? *Journal of Children's Services* 2, (2) 4-14.

Ingoldsby, E., & Shaw, D.S. (2002). Neighborhood contextual factors and the onset and progression of early-starting antisocial pathways. *Clinical Child and Family Psychology Review, 5*, 21-55

Jones, K., Daley, D., Hutchings, J., Bywater, T., & Eames, C. (2008). Efficacy of the Incredible Years Basic Parent Training Programme as an early intervention for children with Conduct Disorder and ADHD: Long-term follow-up. *Child: Care Health and Development 34*, 380-390.

Kane GA, Wood VA, Barlow J (2007) 'Parenting Programmes: A systematic review and synthesis of qualitative research', *Child: Care, Health And Development,* 33 (6), 784 - 793 (0305-1862)

Kaufman J, Birmaher B, Brent D, Rao U (1997) Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (KSADS- PL): initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry,* 36, 980–988

Kelly, Y, Sacker, A, Del Bono, E, Francesconi, M, Marmot, M. (2011). What role for the home learning environment and parenting in reducing the socio-economic gradient in child development? Findings from Millennium Cohort Study. *Archives of Disease in Childhood*, 96, 832-837

Kim, H. K., Capaldi, D. M. Pears, K. C., Kerr, D. C. R., & Owen, L. D. (2009) Intergenerational transmission of internalising and externalising behaviours across three generations: Gender specific pathways. *Criminal Behavior and Mental Health, 19,* 125-141.

Klett-Davies M. (ed.) (2010). *Is parenting a class issue?* London, Family & Parenting Institute.

Kling, Å, Forster, M, Sundell, K, & Melin, L. (2010). A randomized controlled effectiveness trial of parent management training with varying degrees of therapist support. *Behavior Therapy, 41,* 530-542.

Kraemer, H., Wilson, G., Fairburn, C., & Agras, W. (2002). Mediators and moderators of treatment effects. *Archives of General Psychiatry, 59,* 877-883.

Kumpfer, K. et al. (2002). Cultural sensitivity and adaptation in family-based prevention interventions, *Prevention Science.* 3, 241-246.

Lambert P, Sutton A, Abrams K, Jones D (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology* 55, 86–94.

Langton EG, Collishaw S, Goodman R, Pickles A, Maughan B, (2011). An emerging income differential for adolescent emotional problems. *Journal of Child Psychology & Psychiatry*, 52, 1081-1088

Le Couteur, A. & Gardner, F. (2008). Structured interviews and observations. In M. Rutter, D. Bishop, D. Pine, S. Scott, J. Stevenson, E. Taylor & A. Thapar (eds), *Rutter's Child and Adolescent Psychiatry,* 5th Edition. Blackwell.

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly*. Annals of the American Academy of Political and Social Science,* 587, 69–81.

Lorenc, T, Mark Petticrew, M, Welch, V, Tugwell, P (2012). What types of interventions generate inequalities? Evidence from systematic reviews. *Journal of Epidemiology & Community Health* doi:10.1136/jech-2012-201257.

Lundahl B, Risser HJ, Lovejoy MC. (2006). A meta-analysis of parent training: moderators and follow-up effects. *Clinical Psychology Review* 26:86-104.

McGilloway, S, Ni Mhaille, G, Bywater, T, Leckey, Y, Kelly, P, Furlong, M, Comiskey, C, & Donnelly, M. (2012). A parenting intervention for childhood behavioral problems: A randomised controlled trial in disadvantaged community-based settings. *Journal of Consulting and Clinical Psychology, 80*, 116-27

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychological Review, 100*, 674-701.

Moran P., Ghate, D. and van der Merwe, A. (2004). *What works in parenting support?* A review of the international evidence. London: Home Office.

Morch, W-T, Clifford, G, Larsson, B, Rypal, P, Tjeflaat, T, Lurie, J, Drugli, B, Fossum, S, & Reedtz, C. (2004). *The Norwegian Webster-Stratton program 1998-2004*. Norway: University of Trondheim.

Murray J, Farrington DP (2010), Risk factors for conduct disorder and delinquency: key findings from longitudinal studies. *Canadian J Psychiatry* 55, 633-42

The National Research Council and Institute of Medicine (2009). *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities.* Washington DC, The National Academies Press.

Nelder, J. A. & Wedderburn, R. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General).* 135, 370-384.

Noble, M, Wright, G , Smith, G & Dibben, C. (2006). Measuring multiple deprivation at the small-area level. *Environment and Planning A*, 38, 169 -185

O'Brien, B. J. and A. H. Briggs (2002). Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. *Stat Methods Med Res* 11, 455-468.

Office for National Statistics (2004). *Mental health of children and young people in Great Britain.* London*:* ONS.

Odgers C, Caspi A, Broadbent J, et al (2007). Prediction of differential adult health burden by conduct problem subtypes in males. *Arch Gen Psychiatry*. 64, 476-484

O' Neill D, McGilloway S, Donnolly M et al (2012). Cost-effectiveness analysis of Incredible Years parenting program in reducing child health inequalities. *European J Health Economics*, in press

Patel, A, Calam, R & Latham, A. (2011). Intention to attend parenting programmes: does ethnicity make a difference? *Journal of Children's Services*, 6(1), 14

Petrosino et al. (2002). *The effects of Scared Straight & other juvenile awareness programs on delinquency*. *The Cochrane Database of Systematic Reviews* Issue 3. John Wiley & Sons.

Petticrew M, Tugwell P, Kristjansson E, Oliver S, Ueffing E, Welch V. (2011). Damned if you do, damned if you don't: subgroup analysis and equity. *Journal of Epidemiology and Community Health.* doi:10.1136/jech.2010.121095

Piquero A, Shepherd I, Shepherd J, Farrington D. (2011). Impact of Offender Trajectories on Health: Disability, Hospitalization, and Death by Middle Age in the Cambridge Study in Delinquent Development. *Criminal Behaviour & Mental Health 21*, 189-201.

Prinz, R.J., Sanders, M.R., Shapiro, et al. (2009). Population-based prevention of child maltreatment: the US Triple P system population trial. *Prevention Science, 10*, 1-12.

Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* 128 , 301-323.

Reid, M., Webster-Stratton, C., & Beauchaine, T. (2001). Parent Training in Head Start: Comparison of Program Response Among African American, Asian American, Caucasian, and Hispanic Mothers. *Prevention Science,* 2, 209-227.

Reyno S, McGrath, P. (2006). Predictors of parent training efficacy for child externalizing behavior problems – a meta-analytic review. *Journal of Child Psychology & Psychiatry, 47*, 99-111

Riley, R, Lambert, P & Abo-Zaid, G.(2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221

Robinson, E., Eyberg, S., & Ross, A. (1980). The standardization of an inventory of child conduct problem behaviours. *Journal of Clinical Child Psychology*, 9, 22–28.

Rutter, M. (2006). Is Sure Start an Effective Preventive Intervention? *Child and Adolescent Mental Health* 11, 135–141.

Sainsbury Centre for Mental Health (2009). The chance of a lifetime: Preventing early conduct problems and reducing crime. London: Sainsbury Centre for Mental Health.

Sampson, R J. & Morenoff, J D. (2004). Durable Inequality: Spatial Dynamics, Social Processes, and the Persistence of Poverty in Chicago Neighborhoods. In S. Bowles, S. Durlauf, & K. Hoff (Eds.), *Poverty Traps.* Princeton, NJ: Princeton University Press.

Scott S (2010). National dissemination of effective parenting programmes to improve child outcomes. *British Journal of Psychiatry* 196, 1-3

Scott, S (2012) Parenting quality and children's mental health: biological mechanisms and psychological interventions. *Current Opinion in Psychiatry* 25:301–306

Scott S, Knapp M, Henderson J, Maughan B. (2001). Financial cost of social exclusion : follow up study of antisocial children into adulthood. *British Medical Journal* 323, 191-4.

Scott S, O'Connor T, Futh A, Price J, Matias C & Doolan M. (2010b). Impact of a parenting program in a high-risk, multi-ethnic community: The PALS trial *Journal of Child Psychology & Psychiatry* 51, 1331-1341.

Scott S, Sylva K, Doolan M, Price J, Jacobs B, Crook C, et al. (2010a). Randomised controlled trial of parent groups for child antisocial behaviour targeting multiple risk factors: the SPOKES project. *Journal of Child Psychology & Psychiatry* 51:48-57.

Schepman, K., Collishaw, S., Gardner, F., Maughan, B, & Pickles, A. (2011). Do changes in parent mental health explain trends in youth emotional problems? *Social Science and Medicine,* 73(2), 293-300.

Serketich, W. J., & Dumas, J. E. (1996). The effectiveness of behavioral parent training to modify antisocial behavior in children: A meta-analysis. *Behavior Therapy*, *27*, 171-186.

Shadish, W. R., & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason we don't let dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology,* 59, 883–893.

Sterne, J. Egger, M. & Moher, D. (2008). Addressing reporting biases. In Higgins, J, Green, S. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley, p. 297–334

Taylor, E., Schachar, R., Thorley, G., & Weiselberg, M. (1986). Conduct disorder and hyperactivity -1 Separating of hyperactivity and anti-social conduct in British child psychiatric patients. *British Journal of Psychiatry*, 149, 760–767.

Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, et al. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health & Quality of Life Outcomes* 5: 63

Thompson, S & Higgins, J (2005). Can meta-analysis help target interventions at individuals most likely to benefit? *Lancet* 365: 341e6.

UNODC (2009). *Guide to implementing family skills training programmes for drug abuse prevention.* New York: United Nations.

Utting D. (2010). *Assessing and meeting the need for parenting support services*. London, Family & Parenting Institute.

van Hout, B. & Al, M. et al. (1994). Costs, effects and C/E-ratios alongside a clinical trial. *Health Econ* 3, 309-319.

Waylen A, Stallard N, Stewart-Brown S (2008). Parenting and health in mid-childhood: a longitudinal study. *European Journal of Public Health*, 18(3):300-5.

Webster-Stratton, C & Reid, J (2010a). The Incredible Years: In: J. Weisz & A. Kazdin, *Evidence-based Psychotherapies* (2nd edn), NY: Guilford

Webster-Stratton, C. (2009). Affirming Diversity: Multi-Cultural Collaboration to Deliver the Incredible Years Parent Programs. *International Journal of Child Health & Human Development*, 2, 17–32.

Webster-Stratton C. & Reid, J. (2010b). Adapting the Incredible Years, an evidence-based parenting programme, for families involved in the child welfare system. *Journal of Children's Services, 5*, 25-42.

Webster-Stratton, C., Reid, J., & Beauchaine, T.P. (2011). Combining parent and child training for young children with ADHD. *Journal of Clinical Child and Adolescent Psychology, 40*, 191-203.

Weisz JR, Weiss B, Han SS, et al (1995). Effects of psychotherapy with children and adolescents revisited: a meta-analysis of treatment outcome studies. *Psychological Bulletin,117*, 450-468

Welch V, Petticrew M, Ueffing E, Benkhalti Jandu M, Brand K, et al. (2012) Does Consideration and Assessment of Effects on Health Equity Affect the Conclusions of Systematic Reviews? A Methodology Study. *PLoS ONE,* 7: e31360.

White IR, Horton N, Carpenter J, Pocock SJ. (2011). Strategy for intention to treat analysis in randomised trials with missing outcome data. British Medical Journal.342:d40.

WHO (2010). *Violence prevention: The evidence.* Geneva: WHO.

Willan, A. R. (2001). Analysis, sample size, and power for estimating incremental net health benefit from clinical trial data. *Controlled Clinical Trials,* 22, 228-237.