National Institute for Health and Care Excellence

Final

Attention deficit hyperactivity disorder (update)

Appendix 1: Cost-effectiveness analysis: What is the cost effectiveness of parent training compared to no treatment for children with ADHD?

NICE guideline CG72

Economic analysis report

March 2018

Final

This guideline was developed by the National Guideline Centre

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and, where appropriate, their carer or guardian.

Local commissioners and providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Contents

1 Cost-effectiveness analysis: What is the cost effectiveness of parent training compared to no treatment for children with ADHD?

1.1 Introduction

The objective of this model is to assess the cost effectiveness of parent training for children with ADHD. In its most pure form, parent training aims to teach the parents of children with ADHD behavioural techniques to encourage the development of coping strategies for managing the behavioural disturbance of ADHD.

Other components that could be part of the training include; the child also being present or attending some behavioural training themselves, or the child's teacher being involved to allow the techniques to be applied during school.

A model on parent training compared to no treatment was constructed for the previous ADHD guideline; therefore this model can be considered an update. Being updated is; the model structure, the treatment effects, the utility values, and the costs and resource components involved in providing the intervention.

1.2 Methods

1.2.1 Model overview

1.2.1.1 Comparators

Being compared is parent training compared to no treatment. No treatment implies no parent training is being offered to the control group. However, in all of the studies included for treatment effect, a proportion of the children are on another current treatment or on 'treatment as usual', which most often is medication but could also be a number of other interventions. Trials with a completely drug naïve population were not available. Therefore the baseline response is assumed to be the underlying response rate of a general population, whereby some children are on treatment and some children are not.

Ideally we would have liked to compare parent training with no parent training in a drug naïve population to be able to answer the question of whether parent training is cost effective as a first line intervention, but this was not possible due to lack of data.

Due to heterogeneity in the studies, a number of base case scenarios have been undertaken, most of which model the treatment effect from an individual study at a time. For this reason, the intervention can vary in terms of the number of sessions, the length of sessions, and who the intervention is targeted at (parents only, parents and children, parents and children and teachers). With resources involved naturally being higher for interventions that involve separate sessions for different audiences.

1.2.1.2 Population

The population is children with ADHD, ranging from age 5 to 14.

1.2.1.3 Time horizon, perspective, discount rates used

The time horizon is 12 months, as little is known about the long term impact of treatment and progression of ADHD. Therefore no discounting is necessary.

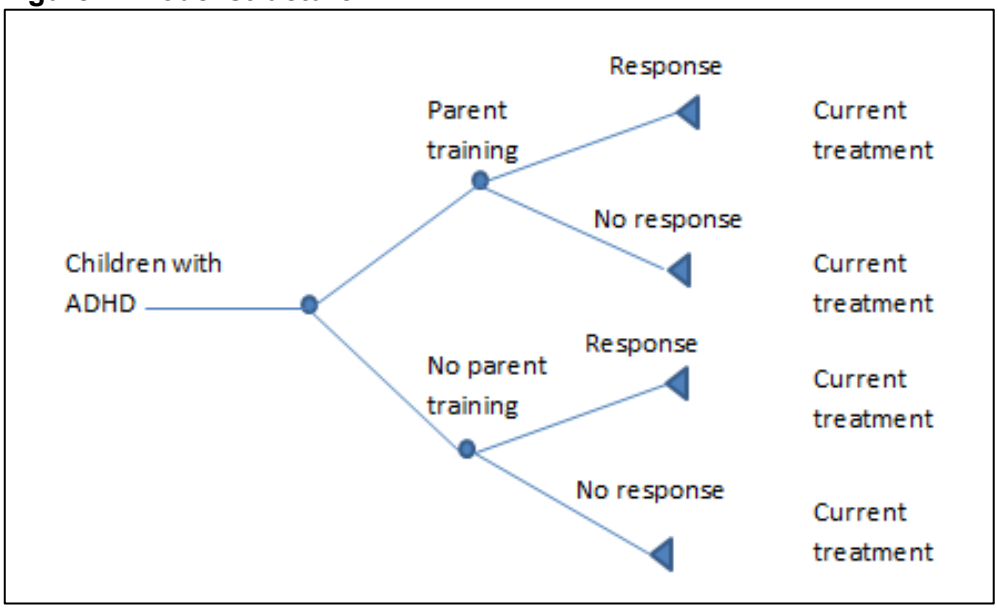
The perspective is that of the NHS and personal social services.

1.2.2 Approach to modelling

The model is a decision tree model, incorporating the cost of the intervention, quality of life, and the treatment effect. The treatment effect is a dichotomous outcome of the proportion of children responding to treatment in the intervention and no treatment arms. It was necessary to use dichotomous outcomes because this was the only way to link to quality of life outcomes, which from the current literature base can only be applied to responders or non-responders.

1.2.2.1 Model structure

Figure 1: Model structure



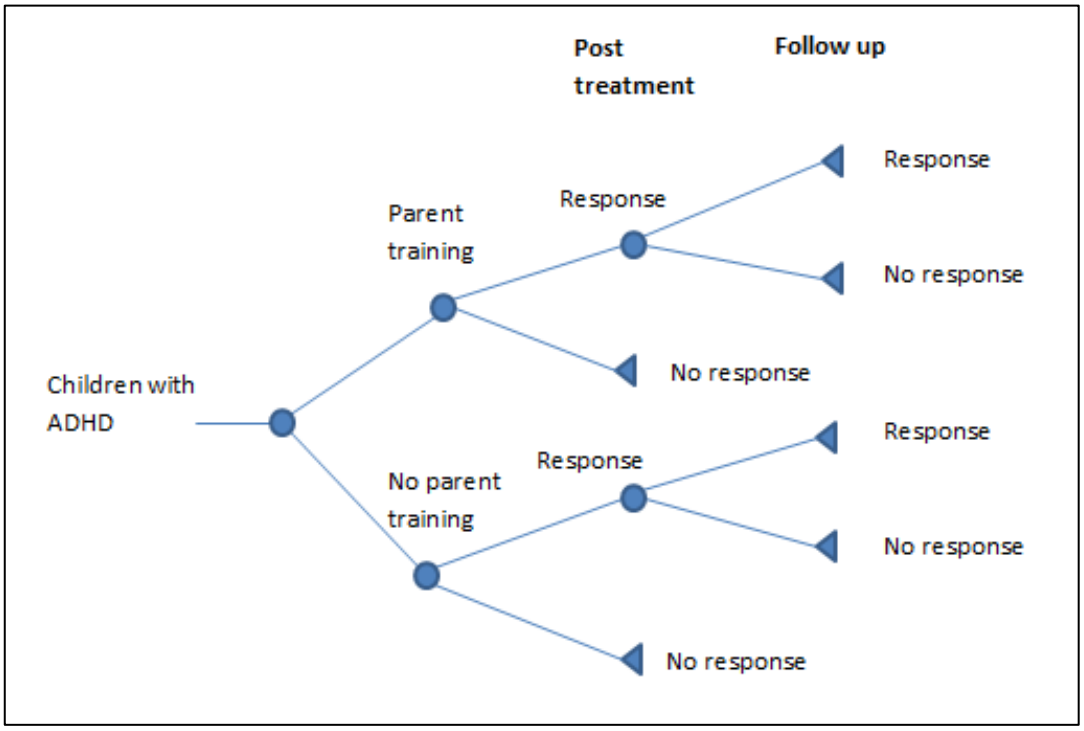
In the previous guideline model, there were also some booster sessions for responders, however the committee felt that this would not be a necessary part of the structure because firstly; if anything it should be non-responders that have the booster sessions, and secondly the booster sessions were not considered to be a routine part of how parent training was offered in the NHS.

There have been no assumptions made about any treatments that patients might go on to following non-response to parent training. Assuming that parent training may be near to the beginning of the treatment pathway, it is probably reasonable to assume that drugs may be tried after a behavioural therapy has failed. However as there is a baseline assumed in the comparator arm that could consist of a number of treatments, adding further treatments into the model after parent training means the intervention would become whether parent training should be first line or not. Ideally a model for ADHD would compare all treatment options individually and in sequences to identify what the most cost effective treatment pathway as a whole might be, but data is not available to be able to do this. So the purpose of this model is

to see if the addition of parent training to what might be considered current practice (as some of the children in the trials are on other treatments) is cost effective, without muddying this with further lines of treatment.

It is also important to note that in order to utilise various data sources on treatment effect, but not being able to combine them all because of heterogeneity, that in one analysis in particular as part of the base case the structure is slightly different because if there were multiple timepoints of data recorded then the structure of the decision tree becomes that of Figure 2, where response at later points in time have been treated independently of response post treatment.

Figure 2: Model structure if include multiple timepoints of data



1.2.2.2 Uncertainty

The model was built probabilistically to take account of the uncertainty around input parameter point estimates. A probability distribution was defined for certain input parameters. When the model was run, a value for each input was randomly selected simultaneously from its respective probability distribution; mean QALYs were calculated using these values, (but not mean costs as these were averages from national sources and therefore it wasn't necessary to subject them to probabilistic sensitivity analysis). The model was run repeatedly – 10,000 times for the base case scenarios – and results were summarised.

The way in which distributions are defined reflects the nature of the data, so for example utilities were given a beta distribution, which is bounded by 0 and 1, reflecting that a quality of life weighting will not be outside this range. All of the variables that were probabilistic in the model and their distributional parameters are detailed in Table 1 and in the relevant input summary tables in Section 1.2.3.1. Probability distributions in the analysis were parameterised using error estimates from data sources, or assumptions.

Table 1: Description of the type and properties of distributions used in the probabilistic sensitivity analysis

Parameter	Type of distribution	Properties of distribution
Baseline risk	Beta	Bounded between 0 and 1. Derived from mean of a domain or total quality of life score and it's the sample size, using the method of moments. Alpha and Beta values were calculated as follows: Alpha = mean×N Beta = N-Alpha
Treatment group risk	Lognormal	Lognormal Bounded at 0. Derived from log (of the RR) and standard error. μ = ln(RR) SD(μ) = (ln[UpperCI] – ln[lowerCI])/1.96*2
Utility	Beta (b)	Bounded between 0 and 1. Derived from mean of a domain or total quality of life score and it's the sample size, using the method of moments. Alpha and Beta values were calculated as follows: Alpha = mean×N Beta = N-Alpha
Incremental utility	Gamma	Gamma distribution: Bounded at 0, positively skewed. Derived from mean and its standard error (a). Alpha and Beta values were calculated as follows: Alpha = (mean/SE)2 Beta = SE2/Mean

(a) The standard error was derived for this from the p-value for the difference between responders and non-responders, the source of this method can be found here: http://handbook.cochrane.org/chapter_7/7_7_3_3_obtaining_standard_deviations_from_standard_errors.htm

(b) Responder utility was incorporated into the probabilistic analysis using a beta distribution. This is bounded by 0 and 1 – although utility can technically go below 0 the values being used here are far from 0 and so this was considered reasonable. This was parameterised using the reported n number from the study group. While technically this approach is for dichotomous data given that no estimate of variability was reported in the study the only other approach would be to make an assumption about variability. Using the n number to parameterise a beta distribution will at least reflect that variability will be lower when the study population is higher and so was considered preferable to assuming a SE.

The following variables were left deterministic (that is, they were not varied in the probabilistic analysis):

- the cost-effectiveness threshold (which was deemed to be fixed by NICE),
- the resource, including time and cost of staff, required to implement each strategy (assumed to be fixed according to national pay scales and programme content)

In addition, various threshold analyses and deterministic sensitivity analyses were undertaken to test the robustness of model assumptions. In these, one or more inputs were changed and the analysis rerun to evaluate the impact on results and whether conclusions on which intervention should be recommended would change.

1.2.3 Model inputs

1.2.3.1 Summary table of model inputs

Model inputs were based on clinical evidence identified in the systematic review undertaken for the guideline, supplemented by additional data sources as required. Model inputs were validated with clinical members of the Committee. A summary of the model inputs used in the

base-case (primary) analysis is provided in Table 3 below. More details about sources, calculations and rationale for selection can be found in the sections following this summary table.

Table 2: Summary of base-case model inputs

Input	Data	Source
Population	Children with ADHD (age 6-14)	Studies informing treatment effect
Time horizon	12 months	
Length of treatment	10 weeks	Average of clinical review studies and GC opinion
Baseline effect	Probability of response	Various trials from the guideline clinical review
Treatment effect	Probability of response	Various trials from the guideline clinical review

A variety of analyses were run as part of the base case because the studies that had the relevant interventions and outcomes that were identified as potentially usable to inform a model, turned out to be quite different studies that had a lot of heterogeneity, and so most of these were analysed separately.

Table 3: Overview of parameters and parameter distributions used in the model

Parameter description	Point estimate	Probability distribution	Distribution parameters	Source
Effectiveness of comparators in base case scenarios				
Analysis: Base case_CHACKO and HANDEN pooled				
Baseline response probability	0.127	Beta	Alpha = 9 Beta = 62	Crude summation of responders divided by total number of people in baseline arms.
Parent training response probability (at 9 weeks)	0.217	Lognormal	LN mean = -1.283(a) SE = 0.45	Derived from the odds ratio of the response in the treatment group versus the baseline group from pooling Chacko 2009 [3] and Handen 2015 [10].
Analysis: Base case_CHACKO				
Baseline response probability	0.075	Beta	Alpha = 3 Beta = 37	Chacko 2009 [3]
Parent training response probability (at 9 weeks)	0.15	Lognormal	LN mean = -1.733(a) SE = 0.676	Derived from the odds ratio of the response in the treatment group versus the baseline group from Chacko 2009 [3].
Analysis: Base case_HANDEN				
Baseline response probability	0.194	Beta	Alpha = 6 Beta =25	Handen 2015 [10]
Parent training response probability (at 9 weeks)	0.290	Lognormal	LN mean = -0.897 (a) SE = 0.604	Derived from the odds ratio of the response in the treatment group versus the baseline

Parameter description	Point estimate	Probability distribution	Distribution parameters	Source
				group from Handen 2015 [10]
Analysis: Base case_PFIFFNER				
Baseline response probability (post treatment)	0.66	Beta	Alpha = 20 Beta =10	Pfiffner 2007 [19]
Parent training response probability (post treatment – 12 weeks)	0.98	Lognormal	LN mean = 4.28 (a) SE = 1.474	Derived from the odds ratio of the response in the treatment group versus the baseline group (short term) from Pfiffner 2007 [19]
Baseline response probability (6 months follow up)	0.40	Beta	Alpha = 12 Beta =18	Pfiffner 2007 [19]
Parent training response probability (6 months follow up)	0.56	Lognormal	LN mean = 0.247 (a) SE = 0.53	Derived from the odds ratio of the response in the treatment group versus the baseline group (follow up) from Pfiffner 2007 [19]
Analysis: Base case_OSTBERG				
Baseline response probability	0.56	Beta	Alpha = 19 Beta =15	Ostberg 2012 [17]
Parent training response probability (at 24 weeks)	0.67	Lognormal	LN mean = 0.693 (a) SE = 0.494	Derived from the odds ratio of the response in the treatment group versus the baseline group from Ostberg 2012 [17].
SA1 (using studies with behavioural outcomes)				
Baseline response probability	0.22	Beta		Crude summation of responders divided by total number of people in baseline arms of Chacko 2009[3] and Fabiano 2012[9].
Parent training response probability (at 9 weeks)	0.36	Lognormal		Derived from the odds ratio of the response in the treatment group versus the baseline group from pooling Chacko 2009[3] and Fabiano 2012[9]..
Cost (£)				
Clinical psychologist (Band 8a)	£62 per hour			PSSRU 2016 [6]
Assistant (Band 4)	£30 per hour			PSSRU 2016 [6]
Consultant Psychiatrist (a)	£208 per hour of patient contact	NA		Cost components (excluding qualifications) that feed into cost per hour, and total hours worked, are from

Parameter description	Point estimate	Probability distribution	Distribution parameters	Source
				PSSRU 2016. However the 2016 cost was applied to the ratio of direct to indirect patient related activity that was last published in PSSRU 2013 (1:0.95) [5], to derive the cost per hour of patient contact. As for every 1 hour spent with a patient, an additional 95% of 1 hour is spent on indirect activities to do with that patient.
Utilities				
Responder utility	0.83	Beta	Alpha = 489.7 Beta = 100.3	Van Der Kolk 2014 [23]
Non-responder utility	0.74	Beta	Alpha = 436.6 Beta = 153.4	Van Der Kolk 2014 [23]
Utility gain from responder over non-responder	0.09	Gamma	Alpha = 10.94 Beta = 0.008	Difference between responder and non-responder utility

(a) Log odds risk in the treatment group.

1.2.3.2 Initial cohort settings

The intervention is parent training which consists of a certain number of sessions per week and can involve training for only parents or for teachers and children as well (the exact resource use involved in the interventions is dependent on the base case scenario). The intervention is provided by a clinical psychologist, and an assistant. It is a group therapy with 10 sets of parents/families per group.

The quality of life of responders is assumed to increase linearly over the treatment period to reach the quality of life of a responder from that of a non-responder, in order to capture that the intervention is likely to have an effect over time rather than an immediate effect. Following the end of the intervention, the utility of responders is assumed to remain static (remain responding).

1.2.3.3 Baseline event rates

Baseline event rates in the model vary depending on which analyses from the base case scenarios are being run. See Table 3 for more detail on the inputs used for each analysis. This is because attempting to pool them identified a lot of heterogeneity because of differences between the studies. More on this is explained in the next section below on relative treatment effects.

The baselines are interpreted to be the underlying population response rates in the general population in which some patients would be on drugs/some kind of current care. Therefore in each analysis the response rate is assumed to be present from the beginning of the analysis, and does not increase linearly over time as the intervention response does. Conversely it is also not assumed to decrease - where to no information was available on this.

One particular study (Pfiffner 2007)[19] had data from two timepoints available. This means it also had effectiveness data for the baseline arm for two timepoints. Rather than use only the baseline of the earlier timepoint and assume this was constant, the baseline from the follow up timeframe was also used because the response probability at follow up was lower than at post treatment, and therefore it seemed reasonable that there may well be some deterioration over time, perhaps from children getting used to the baseline treatments and them being less effective over time.

In all the other studies where only the post treatment outcome was reported, then as mentioned above the baseline response rate is assumed to be constant. If the population in the studies were an entirely newly diagnosed population, then it would be reasonable to assume their response level could increase over time because of the provision of information about the condition and how to manage it. But as we know this is not the case because some are already on other treatments, the assumption of a baseline response probability that remains stable was felt to be more appropriate.

1.2.3.4 Relative treatment effects

Data were used from the clinical review that reported dichotomous outcomes, as this remains the only way to link to quality of life values. The clinical review for the guideline only extracted continuous outcomes (unless no continuous outcomes were reported in the study in which case dichotomous outcomes were extracted) because the committee felt that continuous outcomes were most helpful to their decision making. Therefore studies with the relevant intervention were identified from the clinical review and assessed for dichotomous outcomes by the health economist, and these were extracted.

For the base case analysis, 8 studies met the inclusion criteria of assessing the relevant intervention (parent training but this may also include additional components about the child and/or teacher), and also reporting dichotomous outcomes. 2 studies were excluded from the model even though they reported dichotomous outcomes because one was a mostly inattentive subtype population and was felt to be difficult to generalise from (Pfiffner 2014)[18], and the other had an intervention felt to be more psycho-education rather than parent training (Daley 2013)[8]. Information on the remaining 6 studies can be seen below in Table 4.

After discussion with the committee, it became apparent that some studies are reporting dichotomous outcomes of response based on criteria of behavioural scales rather than ADHD symptom scales. The committee felt that behaviour is an important aspect of the condition for which the outcomes should be treated as separate to the ADHD total symptoms outcomes. This led to some studies being separated from the overall 6 and the response to treatment based on behavioural scales would be included in a sensitivity analysis.

Three studies were identified as having total symptoms outcomes at a similar timeframe (9-12 weeks). Attempting to pool these studies identified a large amount of heterogeneity (I2 of 53%), and it could be seen from the data from the studies that they were very different, as Pfiffner for example had a baseline risk and treatment risk that was much higher than that of the other studies. This heterogeneity suggested that it wasn't a good idea to pool these studies together as it would give a large amount of uncertainty that when propagated in the model through the PSA would lead to a large variation in the results. It was therefore decided to keep the studies separate and model each of them separately.

The same could also be said of the studies that were informing the behavioural outcomes sensitivity analysis, as again there was one particular outlier there (Hoath[11]) that had a much higher odds ratio than the others included in the pooling for that outcome. It was decided that as that was a particularly small study, of only 22 children in total, that it would be excluded from the analysis entirely as it was adding little to the meta-analysis and only acting to increase uncertainty (this is still included in Table 4 for information). Therefore 5 studies were

included in the model in total in various analyses, with studies only being pooled where heterogeneity was small.

There are a number of reasons as to why there are differences between the studies that may be causing heterogeneity. Some of these are explored below;

- The proportion of children on medication varies in each study, reflecting that there is a mixed population that effectiveness in the model is being based on, ranging from Pfiffner 2007 in which a very small number of children were on medication (as well as this study being mostly inattentive subtype) to other studies where a much higher proportion were on medication. This leads to uncertainty about how effective the intervention might be in a drug naïve population; as if children are already on medication then perhaps they are less likely to benefit from the parent training. But also by having dissimilar numbers of people on concurrent treatment in the two arms of a trial can impact the relative effectiveness of the intervention and in turn the model results.
- The final column of Table 4 shows the type of intervention that was being provided in terms of who the training was being offered to. So some studies were more intensive than others in terms of providing training to the child and/or teacher as well as the parent(s), all with the aim of improving the outcomes of the child.
- Another difference between the studies and a concern the committee had about the data was the dichotomous outcomes that were being measured and what they were actually measuring. The committee thought it important that dichotomous outcomes that measured total ADHD symptoms were used rather than outcomes on subscales (such as hyperactivity or inattentiveness), as trying to combine these to work out total scores ourselves post analysis would come with additional assumptions and wouldn't be methodologically sound. Following this process of extracting only total symptom outcomes still leaves us with different studies using different scales/measures. For example; the CGI-I (Clinical Global Impressions – Improvement scale) is a 7 point scale measuring improvement (it is a relative scale), however improvement could mean improvement in symptoms or improvement in behaviour/function and is therefore capturing more than just improvement in symptoms. The SNAP-IV scale (Swanson, Nolan, and Pelham IV) is an 18 item scale where 9 items measure hyperactivity-impulsivity symptoms and 9 items measure inattentiveness and is therefore more symptom focused than the CGI-I. Similarly the ADHD RS (ADHD rating Scale) is an 18 item rating scale that reflects the DSM-IV criteria (Diagnostic and Statistical Manual of Mental Disorders) and is validated tool for measuring ADHD symptoms of hyperactivity/impulsivity and inattentiveness. The Disruptive Behaviour Disorder rating scale is a 45 question screening measure which is designed to identify symptoms of ADHD, oppositional defiant disorder, or conduct disorder. This could imply that it has questions that are wider than the core symptoms of ADHD as it also captures symptoms of behavioural disorders.

The studies were also reporting different time points; post treatment outcomes from the studies (ranging from 9-12 weeks depending on the length of the treatment in the study), and also follow up outcomes (which were reported as 3 months following the end of the intervention).

Therefore the base case consists of the following analyses based on the studies;

- Base case_CHACKO and HANDEN pooled (9 week intervention)
  - *Chacko 2009* (Disruptive Behaviour Disorders (DBD) rating scale (ADHD symptoms), less than or equal to 1)
  - *Handen 2015* (>=30% decrease on the SNAP and CGI-I<=2)
- Base case_CHACKO (9 week intervention)
  - *Chacko 2009* (Disruptive Behaviour Disorders (DBD) rating scale (ADHD symptoms), less than or equal to 1)
- Base case_HANDEN (9 week intervention)

  - *Handen 2015* (>=30% decrease on the SNAP and CGI-I<=2)
- Base case_PFIFFNER (12 week intervention, and 3 month follow up after treatment ended)
  - *Pfiffner 2007* (CGI-I, based on a description of proportion 'at least slightly improved' in intervention group and 'unchanged or worse' in control group. Slightly improved has been interpreted as =<3 on the CGI-I, and unchanged or worse as >3).
- Base case_OSTBERG (10 week intervention but outcomes reported at 3 month follow up after treatment ended = outcomes reported at around 24 weeks)
  - *Ostberg 2012* (numbers of children who did not meet the criteria for diagnosis on the ADHD-RS)

The Chacko and Handen studies were pooled because a meta-analysis of the two did not reveal any heterogeneity (see forest plots in section 1.5).

Note that although Pfiffner and Ostberg both had similar follow up timeframes, these studies were not pooled together because Pfiffner was felt to be different to the other studies. Although it had similar follow up results to that of ostberg and so wouldn't have changed the results of the Ostberg analysis very much.

The sensitivity analysis using studies that report behavioural outcomes rather than total symptoms outcomes is based on a pooling of Chacko 2009 and Fabiano 2012. As mentioned above the Hoath study was excluded because it was adding heterogeneity and it was also a very small study. The remaining two studies didn't have as much heterogeneity (I2 of 0%) and so it was felt acceptable to pool them. one was a 9 week study and one an 8 week study but outcomes in that analysis were assumed to be at 9 weeks. See section 1.5 for details on heterogeneity for studies that were pooled.

All the studies are looking at group training, however Pfiffner which has additional components has some individual aspects also. See Table 4 below for study details.

Note also that the probability of responding from each study has been included on an intention-to-treat basis. This is particularly relevant for the longer term studies because the denominator from the proportion of responders (if a percentage is reported) or the number of children at follow up (if raw numbers are reported) are based on the follow up population which is less than the initial population that was recruited for the study as some people are always lost to follow up. To continue applying intention to treat principles however, the denominator of the number of people who initially went into the trial has been used. This means that the response probability reported here from the longer term trials may be lower than reported in the trial because we are assuming those that were lost to follow up from baseline were non-responders. This leads to a conservative estimate of effect, but if the intervention is cost effective with these conservative assumptions then it will naturally be more cost effective with an even higher response rate.

The probability of response to parent training (the intervention) was derived by using the log odds ratio of the treatment group compared to the baseline group identified from the analysis (or met-analysis if studies were pooled) in revman software. This log odds ratio was then transformed into a probability through a series of formulaic steps outlined below.

$$\theta = Log\ odds\ of\ p$$

1. Calculate the log odds of the baseline risk

$$Logit(\widetilde{BR}) = Ln\left(\frac{\widetilde{BR}}{1-\widetilde{BR}}\right)$$

(Where $\widetilde{BR}$ denotes the simulated baseline risk.)

2. Add the log odds of the baseline risk to the log odds ratio to get the log odds (θ) of the risk in the treatment group:

$$\widetilde{\theta} = Ln(\widetilde{OR}) + Ln\left(\frac{\widetilde{BR}}{1-\widetilde{BR}}\right)$$

3. Transfer the log odds back to the natural scale to get the risk for the treatment group:

$$\widetilde{TR} = \frac{e^{\theta}}{1+e^{\theta}}$$

If we compare the studies with dichotomous outcomes that are included in the modelling with their respective continuous outcomes that were extracted for the guideline clinical review, then outcomes that had a clinical benefit are;

- on the ADHD total score for Handen 2015 (parent rated, pooling of 2 studies)
- on the CGI-I for Handen 2015
- on an inattention outcome (parent and teacher rated) for Pfiffner 2007
- on the ADHD total score parent rated for Ostberg 2012
- on an inattention outcome (parent and teacher rated) for Pfiffner 2007

For the studies informing response on behavioural/function scales in the model sensitivity analysis, no outcomes had a clinical benefit in the continuous outcomes extracted in the clinical review (for Chacko 2009, Fabiano 2012).

Although for the majority of the outcomes included in the clinical review most individual subscales were not found to have a clinically important benefit, there was benefit on total ADHD symptom scales. This may be because a total ADHD scale, based on the 18 items, will have greater range (in terms of assessing more domains) and variance (of scores) to detect small differences.

It would however be fair to say that for the outcomes that were identified as having clinical benefit from the clinical review, these outcomes are either solely or partly based on studies used in the model for effect.

As we are using dichotomous outcomes, it could be argued that any increase in the number of responders in an intervention arm compared to the control group could be considered clinically effective. What is considered an adequate increase to make the intervention cost effective however will be explored in this model.

Table 4: Dichotomous treatment effects extracted from clinical studies

Study	Population	Intervention	Comparator	Outcomes	Proportion using medication	Classification of intervention based on participants
Base case studies						
Chacko 2009	Children aged 5-12, and their single mothers	**Intervention 1:** Behavioural parent training (held for 2.5 hrs each week. Children participated in a concurrent social skills training programme). Group based. Assumed 9 week program as not mentioned in study. N=40 **Intervention 2:** STEPP program[a] (the same as 1 but also includes enhancements to try and motivate mothers more.) N=40 **Intervention 3:** Behavioural parent training + STEPP. (For the purposes of analysis this group combined the other two, and was not actually an intervention in the trial.)	Waitlist control N=40	Disruptive Behaviour Disorders (DBD) rating scale (ADHD symptoms), less than or equal to 1. **Intervention 1:** 0. 1 **Intervention 2:** 0.2 **Intervention 3:** 0.15 **Comparator:** 0.075 9 weeks	**Intervention 1:** 35% **Intervention 2:** 40% **Intervention 3:** NA **Waitlist control:** 37.5% For children receiving medication parents were asked to maintain the type and dose of medication for the duration of the study, and report any changes in medication.	Parent AND child training.
Handen 2015	Children aged 5-14	Parent training (9 weekly individual meetings of 60-90 minutes). N=31	Placebo N=31	>=30% decrease on the SNAP and CGI-I<=2 **Intervention:** 0.290 **Comparator:** 0.194 9 weeks	'45.3% had received prior treatment for ADHD'.[c]	Parent training
Pfiffner 2007	Children aged 7-11	CLAS[b]. Had 3 components (group based); a) 30 min teacher meeting followed up by 4-5 30 min meetings of teacher, parent, child and therapist over 12 weeks. b)	Control (treatment as usual or waitlist)	12 weeks: CGI-I, based on a description of proportion 'at least slightly	'only two of the children participants were taking	Parent AND child AND teacher training

Study	Population	Intervention	Comparator	Outcomes	Proportion using medication	Classification of intervention based on participants
	Predominantly ADHD-I subtype.	Parent training; parents attended 8-10 90 min sessions and 4-5 family sessions. c) Children attended child group at the same time parents attended parent group. After the 12 week study period, there were monthly meetings until follow up (outcomes are reported post intervention and at follow up. Follow up was around 3 months after treatment ended = around 6 months after treatment began). N=36	N=30	improved' in intervention group and 'unchanged or worse' in control group. Slightly improved has been interpreted as =<3 on the CGI-I, and unchanged or worse as >3. **Intervention:** 1 **Comparator:** 0.66 6 months (follow up 3 months after intervention ended): CGI-I, based on a description of proportion of people "improved or much improved". This has been interpreted as a score of =<2. **Intervention:** 0.63 **Comparator:** 0.40	medication for ADHD when recruited' Families expected to change medication were excluded – implying can be on medication.	
Ostberg 2012	Children aged 10. Only 93% diagnosed with ADHD.	Parent and teacher training (parents meet for 10 weekly 2 hr sessions, and teachers meet for eight sessions). Group based. Outcome time period is 3 months following end of intervention. N=36	Waitlist N=34	numbers of children who did not meet the criteria for diagnosis on the ADHD RS **Intervention:** 0.667 **Comparator:** 0.559 Around 24 weeks	**Intervention:** 86% **Comparator:** 77% Proportions using medication during the intervention time.	Parent AND teacher training

Study	Population	Intervention	Comparator	Outcomes	Proportion using medication	Classification of intervention based on participants
SA: Behaviour scales						
Chacko 2009	As reported earlier in this table	Interventions as reported earlier in this table	As reported earlier in this table	Impairment Rating Scale (any score less than 3) **Intervention 1:** 0.1 **Intervention 2:** 0.2 **Intervention 3:** 0.15 **Comparator:** 0.05	As reported earlier in this table	As reported earlier in this table
Fabiano 2012	Children aged 6-12 years old and their male caregivers	COACHES[d] program (8 week 2 hr behavioural program. First hour of each sessions was fathers learning behavioural techniques, concurrently children played soccer with counsellors, and in the second hour parents and children played soccer together.). Group based. N=28	Waitlist N=27	ECBI intensity score of below 60 **Intervention:** 0.62 **Comparator:** 0.48 8 weeks	54% in each group	Parent AND child training
Hoath 2002	Aged 5-9 years	Enhanced Triple P parenting program. (Attended five 2 hr weekly group sessions. Also had weekly telephone conversation starting the week of the 5th group session). Assumed outcomes are at post intervention time frame of 12 weeks. N=10	Waitlist N=11	Criteria for a positive response was defined as "requiring clinically reliable change on the ECBI intensity score in addition to maintenance within the normal range or clinically reliable change on at least one parenting variable on the Parenting Scale" **Intervention:** 0.77 **Comparator:** 0.18 12 weeks	**Intervention:** 80% **Comparator:** 63%	Parent training

Note that the outcome response rates reported may not be the same as those reported in table 3, because those in this table are the crude estimates from the study, but the intervention response rates in table 3 are derived by applying the odds ratio from Revman to the baseline response rate.
STEPP = Strategies to Enhance Positive parenting
(a) CLAS = Child Life and Attention Skills
(b) Note that there could also be a placebo effect here.
(c) COACHES = Coaching Our Acting-out Children: Heightening Essential Skills

A graphical representation of how the effect would look over the time horizon in each base case analysis can be seen below.

The linear parts of the graph represent that response wouldn't occur immediately and this is applied in the model by assuming the increase in utility from non-response to response for the responders happens linearly (so the additional probability of response above baseline for the aforementioned time periods are multiplied by; utility gain*0.5).

Figure 3: Treatment effect over model time horizon; base case_CHACKO and HANDEN

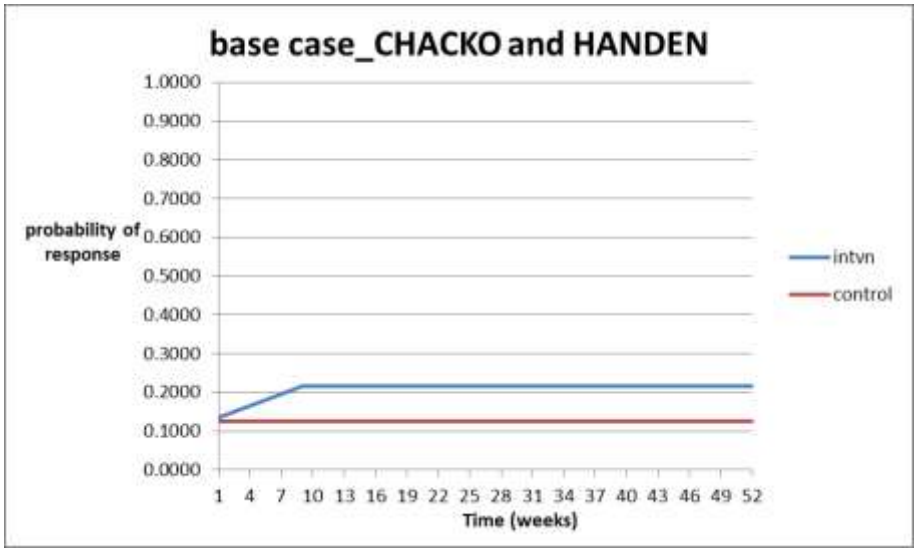


Figure 4: Treatment effect over model time horizon; base case_CHACKO

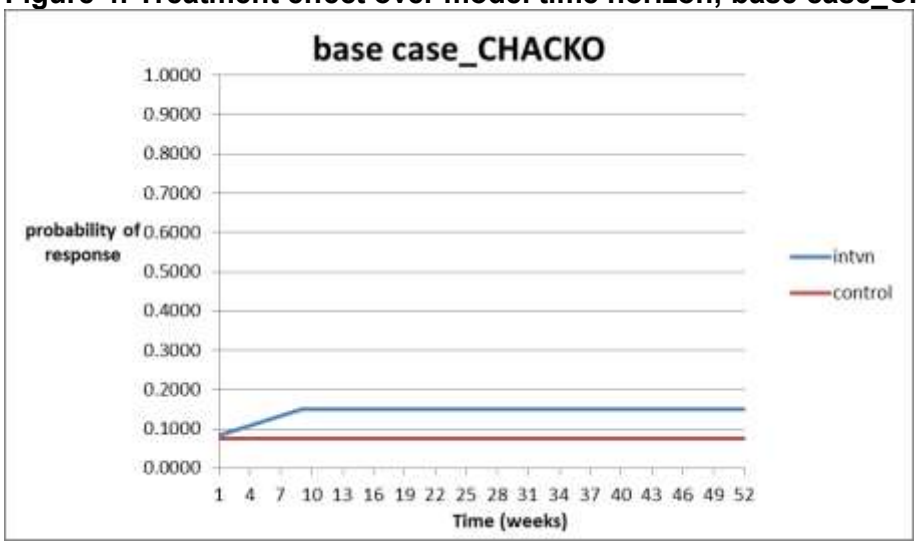


Figure 5: Treatment effect over model time horizon; base case_HANDEN

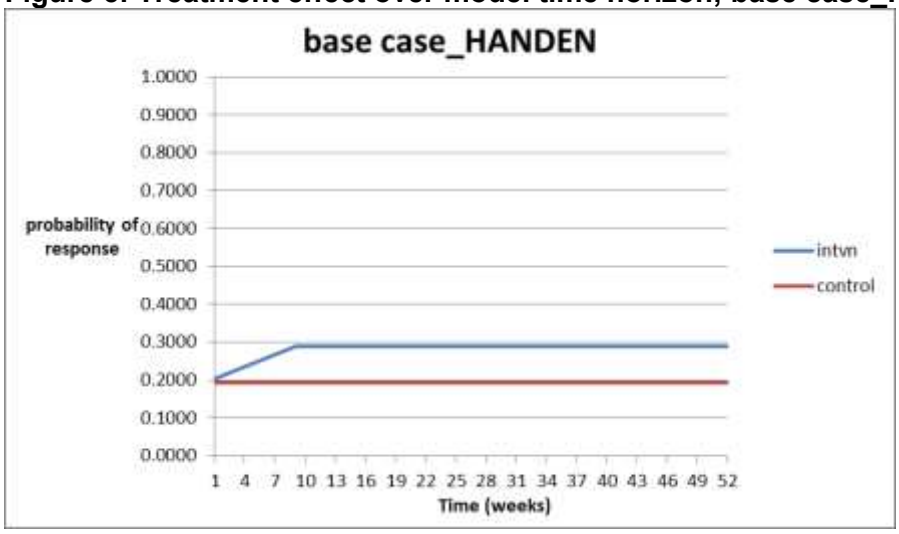


Figure 6: Treatment effect over model time horizon; base case_PFIFFNER

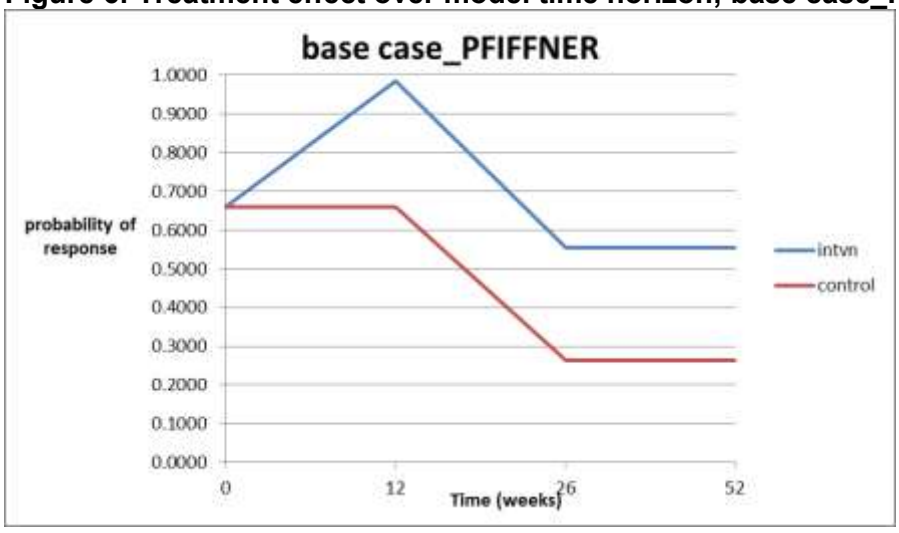
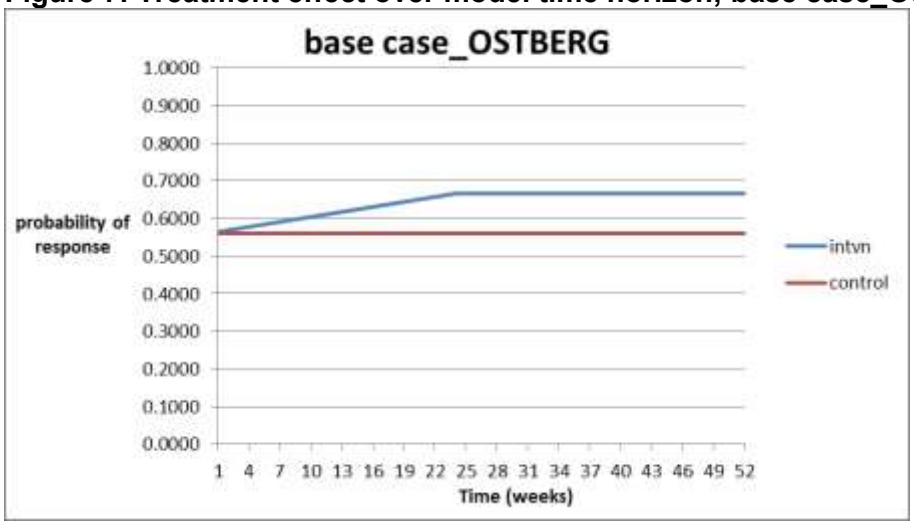


Figure 7: Treatment effect over model time horizon; base case_OSTBERG



1.2.3.5 Utilities

A systematic literature review was conducted to identify quality of life values (this searched for new data following that last guideline). A total of 5 potentially useful studies were identified for children that either used a generic quality of life measure or described vignettes and used methods to elicit preferences directly such as standard gamble or time trade-off. 4 abstracts were also identified as being potentially relevant and these can also be seen in Table 5 below.

Table 5: Utilities for children identified from systematic search

Study	Detail	Utilities
Published studies		
Van der Kolk (2014a) [23]	Conducted a survey in the Netherlands to collect data on QoL of children with ADHD that were on drug treatment and their parents. Used EQ-5D (and Kidscreen 10). Grouped people into responders and non-responders by coming up with descriptions for those groups prior to the study. - A responder was a child that was compliant (taking the prescribed dose) and functioning fine. - A non-responder was not using the prescribed dose and had some problems functioning. Total sample of 428 responders and 190 non responders.	Scores reported (using UK tariff) are; QoI of children from parents: 0.80 (entire sample), 0.83 (responder), 0.74 (non responder). The groups were also subgrouped by factors like age (8-12 and 13-18).
Carroll 2009[2]	Utilities calculated for a wide range of health states in the paediatric population, but elicited from parents. 2 states of ADHD were also included (mild and severe ADHD). Elicited using the health state descriptions with Standard Gamble and Time-Trade-Off methods. States were assessed around 400 times.	SG mean values: 0.94 (mild ADHD), 0.92 (Severe ADHD) TTO mean values: 0.93 (mild ADHD),

		0.90 (Severe ADHD)
Lloyd 2011 [15] Population: 20 children ages 11-16 were used for qualitative interviews to develop the ADHD health states. These states were then rated by 100 members of the UK public.	This study involved a series of steps: Step 1: health states were developed by interviewing young people with ADHD and their parents, and analysis of baseline clinical trial data (a methylphenidate trial was analysed and participants were grouped according to their baseline CGI-S level, and the frequency of responses to individual questions on the ADHD-RS and AIM-C helped with identifying text to include in the health state description). Step 2: health states were valued by 100 members of the UK public using the TTO method. States included; normal, borderline to mildly ill, moderately to markedly ill, severely ill (based on the CGI-S but reduced into fewer states). Step 3: Mapping the CGI-S to the CGI-I; this was based on data from 2 clinical trials. CGI-S values at follow up were estimated using linear regression of CGI-S scores on ADHD-RS at baseline (so baseline relationship between the two scores used to predict follow up relationship). The predicted CGI-S scores at follow up were then tabulated by whether or not patients had responded to treatment, where response to treatment was defined as having achieved either the top two or three scores on the CGI-I at the last visit. Mean TTO utilities for responders/non responders were calculated by a weighted average of the CGI-S score frequencies.	TTO scores for the CGI-S states; 0.839 (normal), 0.787 (borderline to moderately ill), 0.578 (moderately to markedly ill), 0.444 (severely ill) TTO scores by classifying responder as >2 on CGI-I; 0.82 (last visit responder), 0.70 (last visit non-responder) (also reports first visit by responder or not)
Bouwmans 2014[1]	A questionnaire survey was performed amongst parents for children with ADHD, this included the EQ-5D. Dutch EQ-5D proxy version used. Around 740 children were included all together and utilities were broken down by; response or not, how many comorbidities were present, and age.	Overall utility for different age groups; 0.81 (8-18) 0.79 (8-11) 0.83 (12-18)
Kandemir 2014[13]	76 children aged 7-16 with ADHD who had been referred to a psychiatrist, these were matched to 59 controls. SF-36 was completed to derive *QoL of the parents*.	
Abstracts		
Van der Kolk 2013[20]	*Children and adolescents aged 8-18 years and their parents. 618 questionnaires to study QoL in children and adolescents with ADHD and their parents, with a focus on compliance.* *Used EQ-5D.*	*Average EQ-5D:* *0.80 (compliant = 0.83, non-compliant = 0.74)*
Van der Kolk 2011[21]	*Parent of a child aged 6-18 with ADHD.* *Comparing QoL in different states of compliance to medication, in remissions status after medication use, or being naïve to medication.* *Using EQ-5D (proxy version for the children).* *873 returned questionnaires.* *Optimal compliance: proxy EQ-5D = 0.8257, EQ-5D= 0.8331*	

	Suboptimal compliance: proxy EQ-5D = 0.7321, EQ-5D =0.8050 *Medication use stopped: proxy EQ-5D = 0.7635, EQ-5D = 0.8169* *Remission after medication: proxy EQ-5D = 0.8518, EQ-5D = 0.8220* *Medication naïve: proxy EQ-5D = 0.7719, EQ-5D = 0.7899*	
Van der Kolk 2014b[24]	*ADHD sample (618) compared to a control group (704). Mean age 11 years.* *EQ-5D for children and parents:* *EQ-5D children ADHD group = 0.80* *Control group = 0.96* *EQ-5D parents ADHD = 0.83* *Control group = 0.88*	
Hodgkins 2013[12]	*Objective of the study was to quantify the utility gain using HUI2 following treatment with lisdex in children and adolescents with ADHD. Compared to OROS MPH. Utilities were estimated for responders and non-responders regardless of treatment.* *Utility for response based on CGI-I of 1 or 2 = 0.896, no response = 0.838* *Utility for response based on ADHD-RS of >25% = 0.899, no response = 0.809* *Utility for response based on ADHD-RS of >30% = 0.902, no response = 0.814*	

The utilities from Van der Kolk 2014a[23] for a responder and non-responder were utilised in the model because it used the UK EQ-5D tariff, and was a fairly large sample. The NICE guideline manual states; “The EQ-5D is the preferred measure of health-related quality of life in adults, and combines both quantity and health-related quality of life into a single measure of health gain. The value placed on health-related quality of life of people using services (or their carers) should be based on a valuation of public preferences elicited from a representative sample of the UK population.”

The utilities are also reported by the parents of children rather than the children themselves. It should also be noted that the utilities from the study are based on responders and non-responders to medication, and therefore may not be as applicable to behavioural therapy because the different interventions affect ADHD symptoms in different ways. There was however no literature identified that specifically looked at the utility of a population only on nonpharmacological therapies.

Because there has been some debate with the committee about whether a generic quality of life measure such as the EQ-5D is responsive enough to capture the quality of life with someone in ADHD and also whether it is sensitive enough to changes in the condition, some alternative ways of measuring utilities have been used in a sensitivity analysis. The guide to the methods of technology appraisal states that; “The measurement of changes in health-related quality of life should be reported directly from patients and the utility of these changes should be based on public preferences using a choice-based method. The EQ-5D is the preferred measure of health-related quality of life in adults.” There is no empiric evidence identifying that the EQ-5D is not valid in an ADHD population. A quality of life search was undertaken for the guideline and was summarised above (for dates see section B.1.2.2 of evidence report E on non-pharmacological efficacy and adverse events), and therefore if

there was data identifying the lack of validity of the EQ-5D this would have been identified. A separate issue that is well recognised is that the EQ-5D has not been designed to be used in children. In the studies specific to an ADHD population, parents tend to evaluate the quality of life of the child.

As well as the impact on the quality of life of children, ADHD may have a wider impact on the quality of life of those around the children such as parents/families or carers. The same study used for the base case utilities (using the EQ-5D UK tariff) also derived the quality of life of the parents. This found that parents of responders had a utility of 0.83, and parents of children that were non-responders had a utility of 0.81. This is a smaller difference than the incremental utility of children, showing that there may be a small impact on parents from the responsiveness of the child. This is only one study however and doesn't mean that there isn't an association between a child's responsiveness and the quality of life of their families.

1.2.3.6 Resource use and costs

1.2.3.6.1 Resource use of providing intervention

The main and only cost involved in providing the intervention is the time costs of staff. Staff costs were sourced from the PSSRU 2016[6].

The costs required include the costs of the clinician that would be teaching the parent training course. The committee thought that on average this would be someone of Band 8a.

As well as the clinician teaching the course, the committee felt it was important to also include the cost of an assistant. The assistant would help with the administration of contacting and inviting parents, setting up the room, and would also be present during the training to assist particularly because it is a group intervention. This is a new addition that was not part of the cost in the previous model, and has increased the cost of the intervention. It was also mentioned how it may not be a paid member of staff that would be an assistant, but might be a volunteer or a student, however the cost of an assistant has been included because there is still an opportunity cost of their time and in order to make the cost effectiveness estimate conservative.

The costs of staff per hour can be found in Table 3. (£62 per hour for a psychologist and £30 per hour for an assistant)

The resources involved in providing the intervention vary depending on the base case analysis, as most analyses modelled individual studies and therefore the resource use from the study was micro-costed. In other words, each component/resource involved in providing the intervention was identified and a cost attached to it, these were then summed up to derive a total cost of providing the intervention within each analysis.

Details of the costs used for the intervention in each analysis can be found in Table 6, except for the costs of the intervention from the Pfiffner study as that had many components and costs for this are presented in Table 7.

Table 6: Intervention costs

Component	Hours spent (on parents)	Hours spent (on children)	Hours spent (on teachers)	Total	Description
Analysis: Base case_CHACKO and HANDEN pooled (a)					
Clinical Psychologist					
Set up time	9	-	-	£558	1 hour for every session would -- be spent preparing. (GC assumption)

Component	Hours spent (on parents)	Hours spent (on children)	Hours spent (on teachers)	Total	Description
Teaching time	18	-	-	£1,116	9 sessions of 2 hours
Assistant					
Set up time	9	-	-	£270	1 hour for every session would be spent setting up. (GC assumption)
Admin time	18	-	-	£540	The administrative tasks involved would take the same number of hours as there are sessions being provided . (GC assumption) (b)
Attending course	18	-	-	£540	Assistants also attend the course to help out where necessary. (GC assumption)
Total cost				= £3,024 **= £302 per family**	**Total cost of providing the intervention**
Analysis: Base case_CHACKO (c)					
Clinical Psychologist					
Set up time	9	9	-	£540	1 hour for every session would -- be spent preparing. (GC assumption)
Teaching time	22.5	22.5	-	£1,350	9 sessions of 2.5 hours (for parents and children)
Assistant					
Set up time	9	9	-	£540	1 hour for every session would be spent setting up. (GC assumption)
Admin time	22.5	22.5	-	£1,350	The administrative tasks involved would take the same number of hours as there are sessions being provided . (GC assumption) (b) (f)
Attending course	22.5	22.5	-	£1,350	Assistants also attend the course to help out where necessary. (GC assumption)
Total cost				= £7,146 **= £715 per family**	**Total cost of providing the intervention**
Analysis: Base case_HANDEN (d)					
Clinical Psychologist					
Set up time	9	-	-	£558	1 hour for every session would -- be spent preparing. (GC assumption)
Teaching time	13.5	-	-	£837	9 sessions of 1.5 hours.
Assistant					
Set up time	9	-	-	£270	1 hour for every session would be spent setting up. (GC assumption)
Admin time	13.5	-	-	£405	The administrative tasks involved would take the same number of

Component	Hours spent (on parents)	Hours spent (on children)	Hours spent (on teachers)	Total	Description
					hours as there are sessions being provided . (GC assumption) (b)
Attending course	13.5	-	-	£405	Assistants also attend the course to help out where necessary. (GC assumption)
Total cost				= £2,475 **= £248 per family**	**Total cost of providing the intervention**
Analysis: Base case_OSTBERG (e)					
Clinical Psychologist					
Set up time	10	-	8	£1,116	1 hour for every session would -- be spent preparing. (GC assumption)
Teaching time	20	-	16	£2,232	10 sessions of 2 hours for parents and 8 sessions for teachers.
Assistant					
Set up time	10	-	8	£540	1 hour for every session would be spent setting up. (GC assumption)
Admin time	20	-	16	£1,080	The administrative tasks involved would take the same number of hours as there are sessions being provided . (GC assumption) (b)
Attending course	20	-	16	£1,080	Assistants also attend the course to help out where necessary. (GC assumption)
Total cost				= £6,048 **= £605 per family**	**Total cost of providing the intervention**
SA1 (using studies with behavioural outcomes) (c) (g)					
Set up time	9	9	-	£1,116	1 hour for every session would -- be spent preparing. (GC assumption)
Teaching time	18	18	-	£2,232	9 sessions of 2 hours for parents and children.
Set up time	9	9	-	£540	1 hour for every session would be spent setting up. (GC assumption)
Admin time	18	18	-	£1,080	The administrative tasks involved would take the same number of hours as there are sessions being provided . (GC assumption) (b)
Attending course	18	18	-	£1,080	Assistants also attend the course to help out where necessary. (GC assumption)
Total cost				= £6,048	**Total cost of providing the intervention**

Component	Hours spent (on parents)	Hours spent (on children)	Hours spent (on teachers)	Total	Description
				= £605 per family	

(a) Intervention: both 9 weeks, Chacko was 2.5 hours per week and was parent and child training, and Handen was 60-90 minutes per week and parent only. As a midpoint, assumed parent training only, with sessions of 2 hour length.
(b) of contacting parents, inviting them and arranging them to attend the course
(c) parent and child training
(d) parent only training
(e) parent and teacher training
(f) An assumption has been made that resource use would be duplicated e.g. if it is both a parent and child training course then the assumptions about time needed to set up etc would be duplicated and there would be no time saving. This may not necessarily be true as it may not be the case that a course that provides training to more than one type of audience would require twice the resources like twice the admin time and twice the set up time. So costs here can be taken as being conservative estimates.
(g) This analysis is a pooling of two studies that report behavioural outcomes. Chacko 2009 had a 9 week intervention with a 2.5 hour session per week. Fabiano 2012 had n 8 week intervention with 2 hours per week. To take the midpoint of the two it was assumed it was a 9 week intervention with sessions of 2 hour length.

Table 7: Intervention costs – Pfiffner 2007 study

			Psychologist		Assistant			
Component	**Number of meetings**	**Length per meetings**	**Total no. of hours face to face**	**Total hours of prep time**	**Admin time**	**Set up time**	**Attending**	
Teacher meeting (a)	1	30 minutes	0.5	- (c)				
Parent, teacher, and child meeting (a)	5	30 minutes	2.5	2.5 (d)				
Parent meetings	10	90 minutes	15	10 (e)		10 (g)	15	
Family sessions (a)	5	30 minutes (b)	2.5	2.5 (d)				
Child meetings	10	90 minutes	15	10 (e)		10 (g)	15	
Post treatment to follow up meetings (monthly) (a)	3	30 minutes (b)	1.5	1.5 (d)				
Total hours			**37**	**26.5**	**37 (f)**	**20**	**30**	
Total hours of individual components			7	6.5	7	-	-	
Total hours of group components			30	20	30	20	30	
Total cost for individual components			£434	£403	£210			= £1,047 per family

			Psychologist		Assistant			
Component	**Number of meetings**	**Length per meetings**	**Total no. of hours face to face**	**Total hours of prep time**	**Admin time**	**Set up time**	**Attending**	
Total cost for group components			£1,860	£1,240.0	£900	£600	£900	= £5,550/10 = £550 per family
Total cost								**= £1,597 per family**

(a) Assumed to be individual
(b) Not state in study so assumed
(c) Assumed negligible
(d) Assumed 30 minutes per meeting
(e) Assumed 1 hour for every session
(f) Assumed to be same number of hours as there are sessions
(g) Only relevant for the group sessions. Assumed one hour per session

Note that no cost exists in the PSSRU for non-contact time cost for a psychologist, so this has assumed to be the same as the face to face time cost.

Current treatment was not assigned a cost because this is applicable for both the intervention and control and would cancel out.

1.2.3.6.2 Costs associated with resource use

Resource use associated with response or no-response over the time horizon of the model was also included because committee opinion was that non-responders would usually be seen more frequently by a psychiatrist/paediatrician than responders. As the underlying population from the studies was children who were on a mix of concurrent treatments or no treatment, rather than a population that were all on medication for example, then it wasn't felt possible to assume that there would be the same underlying resource use for both arms of the model. As if indeed all the underlying population was on medication, then they would be seen regularly by a clinician anyway, and those sessions would double up as checking on the patients progress with other treatments such as parent training, in other words there would be no duplication of resources specifically because of parent training. But as that is not the case, the GC though it was a fair assumption to conclude that patients would therefore be seen with a frequency based on their response to parent training, because they may not be seeing them at all if they are not already on medication. For that reason psychiatrists (or could also be paediatrician) visits were used to represent resource use associated with response.

Although this may be an over or underestimate of the actual resource use reflected based on the response levels from the trials, because those that are on medication would be seen regularly. It can also be the case that parent training may in fact reduce other resource such as reliance on medication. But as we are not certain of the impact of parent training on resource use, an assumption has been made to try and capture the fact that there could be a difference in downstream resource use between responders and non-responders.

Because for the base case analysis that uses the Pfiffner study there are two timepoints, then someone who is a responder in the shorter term may become a non-responder at the later timepoint, and so they may have a different frequency of meetings in the two time periods, and so the resource use has been split into post treatment to 6 months, and 6

months to 12 months. Responders would be seen every 4 months on average, at month 4 and month 8. Non-responders would be seen around twice as much according to the committee, and so it is assumed this might be at months 4,6,8 and 10. The visits start after the course of treatment has finished and so that is why they wouldn't be seen for the first few months. See table below for the costs of this resource use given the number of consultations mentioned.

Table 8: Costs associated with response

Timeframe	Frequency responders seen	Frequency non-responders seen
Post treatment to 6 months	£208	£416
From 6 months to 12 months	£208	£416

In the models where outcomes are from a single timepoint, the resource use for the two periods is summed as you remain a responder or non-responder following treatment for the remaining time horizon of the model.

In the base case analysis using the Ostberg study, although the intervention is a 10 week intervention, the outcomes are only measured at a follow up time period of 3 months post treatment, which would be around 6 months after the intervention began. Because both responders and non-responders have a consultation at 4 months, and the difference between the resource use only really starts at month 6 onwards, then it is assumed that the same costs would be applied for responders and non-responders as in all the other base case models, as you would know at 6 months whether they responded or not.

There are other factors omitted from the model like the cost of treatment people might change to which has monitoring costs (if a drug) as well as the cost of the intervention, which may be different in the different arms because of the impact of parent training. Those on drugs will be seen at regular intervals anyway so it may be irrelevant what impact parent training has, unless it makes them stop drugs completely, or conversely – prevents them from starting drugs in the first place. There are many factors that we are uncertain of and for that reason could not be included in the model and so assumptions had to be made.

1.2.4 Computations

The model was constructed in Microsoft Excel 2010, and evaluated for a single individual. Cohort simulation was not necessary because of the structure and time horizon of the model.

The patient begins at time zero and has the intervention for a set period of time (depending on the study/ies being used for the analysis). Given the baseline and treatment response probabilities; the proportion of people that are responders to either arm are applied the responder utility linearly over the intervention timeframe to represent a slowly increasing level of benefit (through utility) from that of baseline (non-response utility) to that of a responder. In base case analyses where data from two timepoints are included, the probability of response at the later timepoint will then influence the utility gain between the two timepoints and the response probabilities that are carried forward until the end of the time horizon of the model. This level of response is then assumed to remain until the end of the model.

Total costs and QALYs are the sum of the costs in each arm and QALYs in each arm at the end of the model.

In the probabilistic analysis, only the QALYs are probabilistic because no distributions have been put around the costs, however costs will vary through the fact that response rates are varying and costs associated with response are included in the model, and so costs are also recorded for each simulation. The average cost and QALY for each comparison is taken from

all the simulations, and the probabilistic ICER is calculated by dividing the incremental cost by the incremental QALY for each analysis.

1.2.5 Sensitivity analyses

1. Using the outcomes from studies that measure behavioural outcomes dichotomously in the base case, rather than symptom measures.
2. Using continuous outcomes transformed into dichotomous outcomes.
3. Using data from the under 5's population from the guideline review to assess cost effectiveness in that group.

1.2.6 Model validation

The model was developed in consultation with the Committee; model structure, inputs and results were presented to and discussed with the Committee for clinical validation and interpretation.

The model was systematically checked by the health economist undertaking the analysis; this included inputting null and extreme values and checking that results were plausible given inputs. The model was peer reviewed by a second experienced health economist from the National Guideline Centre; this included systematic checking of many of the model calculations.

1.2.7 Estimation of cost-effectiveness

The widely used cost-effectiveness metric is the incremental cost-effectiveness ratio (ICER). This is calculated by dividing the difference in costs associated with 2 alternatives by the difference in QALYs. The decision rule then applied is that if the ICER falls below a given cost per QALY threshold the result is considered to be cost-effective. If both costs are lower and QALYs are higher the option is said to dominate and an ICER is not calculated.

$$ICER = \frac{Costs(B) - Costs(A)}{QALYs(B) - QALYs(A)}$$ Where: Costs(A) = total costs for option A; QALYs(A) = total QALYs for option A	Cost-effective if: • ICER < Threshold

Results are also presented graphically where total costs and total QALYs for each diagnostic strategy are shown. Comparisons not ruled out by dominance or extended dominance are joined by a line on the graph where the slope represents the incremental cost-effectiveness ratio.

1.2.8 Interpreting results

NICE's report 'Social value judgements: principles for the development of NICE guidance'[16] sets out the principles that Committees should consider when judging whether an intervention offers good value for money. In general, an intervention was considered to be cost-effective if either of the following criteria applied (given that the estimate was considered plausible):

- The intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
- The intervention costs less than £20,000 per quality-adjusted life-year (QALY) gained compared with the next best strategy.

1.3 Results

1.3.1 Base case analyses

1.3.1.1 Analysis: Base case_CHACKO and HANDEN

Probabilistic base case results for this analysis can be found below in Table 9:

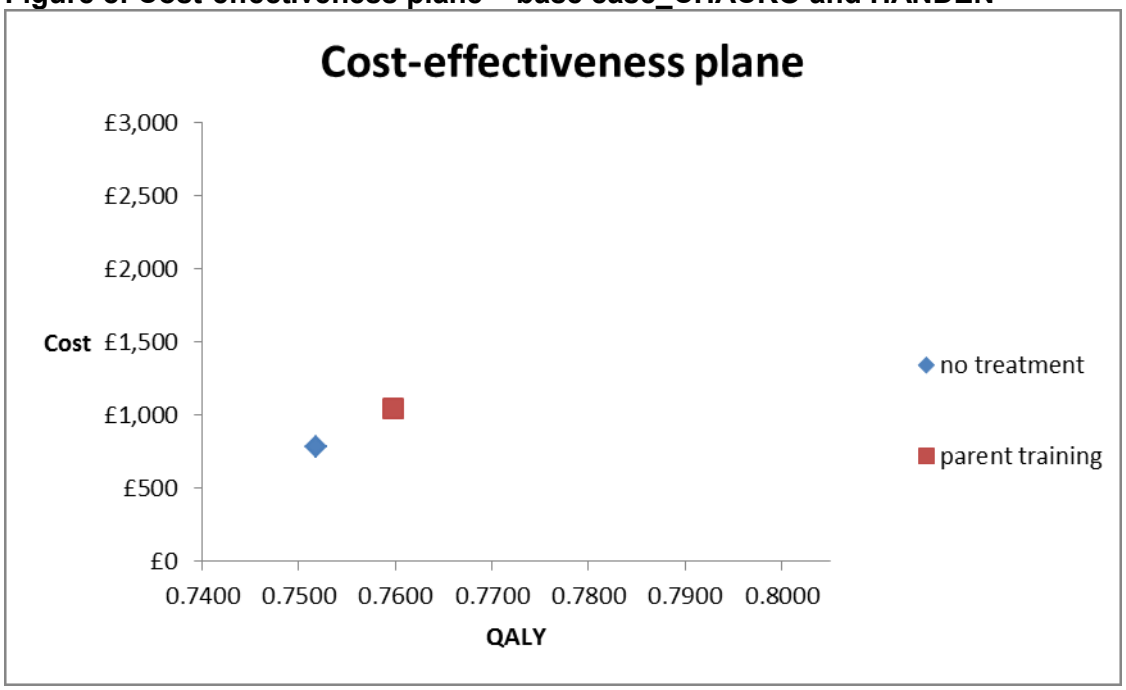
Table 9: Base case results; CHACKO and HANDEN (per person)

	Total cost	Total QALY
Parent training	£1,041	0.7598
No parent training	£779	0.7519
Incrementals	£262	0.0079
ICER	**£33,015**	

The ICER is above the NICE cost per QALY threshold of £20,000, therefore parent training using the two studies pooled here is not considered cost effective.

The cost-effectiveness plane in Figure 8 also shows us the ICER as the gradient of the line that joins no treatment with parent training.

Figure 8: Cost-effectiveness plane – base case_CHACKO and HANDEN



The model is very sensitive to the treatment effect and baseline response rate. As we have a number of base case analyses using different effectiveness and costs, then a variety of ICERS are possible because of the models sensitivity to the inputs. The ICER in this analysis is above the threshold but not too much over £30,000 and this is because the cost of the intervention used is towards the lower end of the scale of because the costs of just parent

training have been assumed, and also the difference in probability between probability of non-response and response is about 9% which impacts incremental QALYs and is bigger than the relative difference from some of the other studies used in the base case.

Given that the underlying population have a significant proportion taking medication, a comment needs to be made on what baseline response rate from the studies would be most similar to what would be expected in practice. This is a difficult question to answer and is highly dependent on both how response is defined and what treatment children - those who are on treatment - would be on.

A threshold analysis was undertaken on cost; keeping all else constant, the cost of the intervention would have to decrease from £302 per person to below £186 to make the intervention cost effective.

Parent training had a 23% probability of being cost effective in this analysis (at £20,000).

1.3.1.2 Analysis: Base case_CHACKO

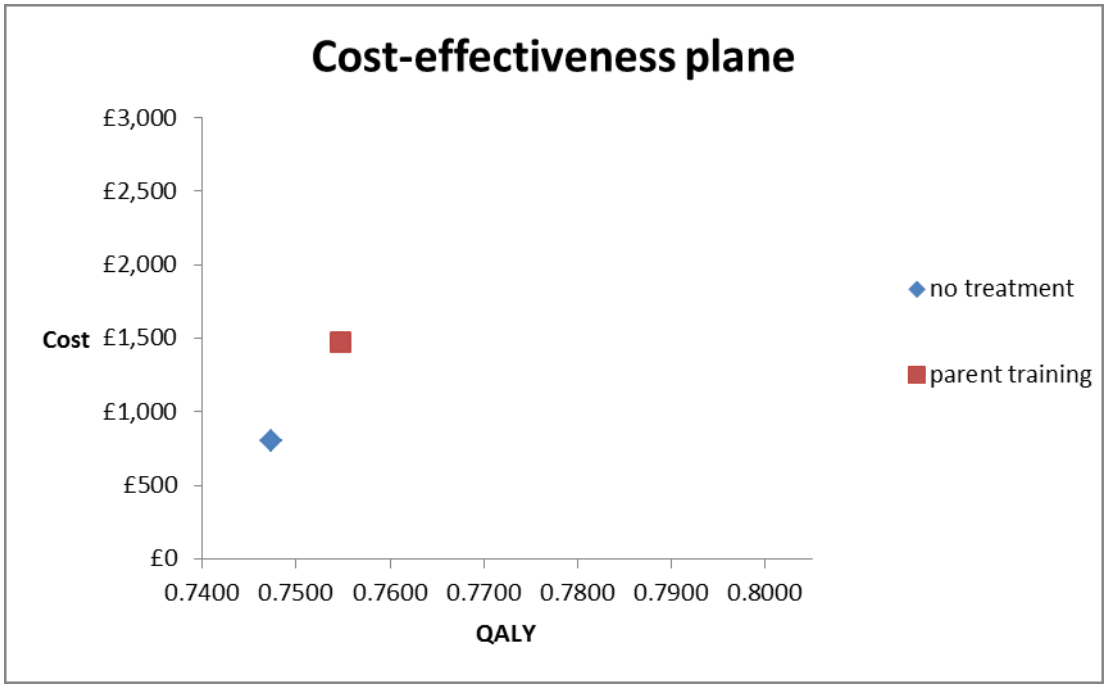
The probabilistic results of this analysis can be found below.

Table 10: Base case results; CHACKO (per person)

	Total cost	Total QALY
Parent training	£1,478	0.7547
No parent training	£800	0.7474
Incrementals	£677	0.0073
ICER	**£92,531**	

In this analysis only the Chacko 2009 study was used for effect. This study was a parent and child training intervention, therefore more resource intensive, and both the baseline and treatment response probabilities are very low, hence why the ICER is high. Parent training only had a probability of 3% of being cost effective in this analysis. The cost effectiveness plane can be seen below.

Figure 9: Cost-effectiveness plane – base case_CHACKO



A threshold analysis on the intervention cost showed that it would have to be below £155 to make the intervention cost effective.

1.3.1.3 Analysis: Base case_HANDEN

The probabilistic results of this analysis can be found below:

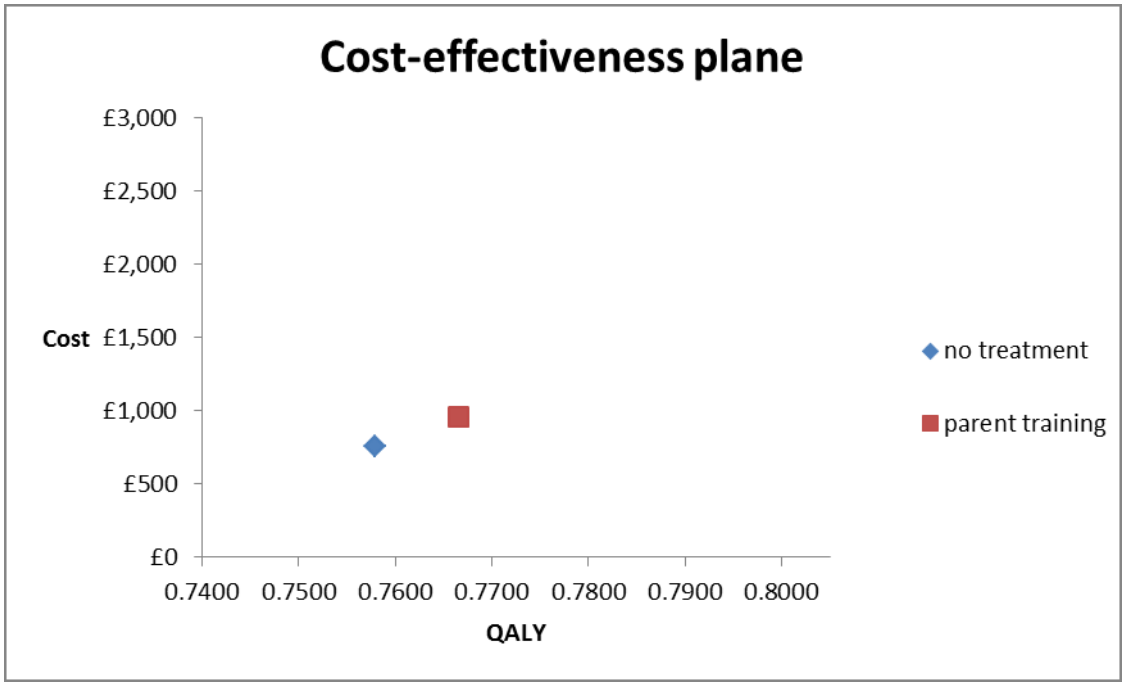
Table 11: Base case results; HANDEN (per person)

	Total cost	Total QALY
Parent training	£955	0.7666
No parent training	£752	0.7579
Incrementals	£203	0.0087
ICER	**£23,393**	

Again we can see that the ICER has fallen significantly from the last analysis. The Handen study was only a parent training study and has the lowest intervention costs of all the studies used because it has 9 sessions of 1.5 hours each. The additional probability of response from the intervention arm is around 10%. Although this isn't very much, we can see that because the cost is low then the intervention is close to being cost effective, even if the intervention is only slightly more effective than the comparator.

The cost effectiveness plane can be seen below. All the cost effectiveness planes have the same axis points so that they can be equally compared. If we were to join the two interventions on the plane below with a line then this would be a line of lower slope than the previous cost effectiveness plane.

Figure 10: Cost-effectiveness plane – base case_HANDEN



A threshold analysis on the intervention cost showed it would have to be lower than £198 to make the intervention cost effective.

Parent training had a probability of being cost effective of 39% at a threshold of £20,000 in the PSA for this analysis.

1.3.1.4 Analysis: Base case_PFIFFNER

The probabilistic results of this analysis can be found below:

Table 12: Base case results; PFIFFNER (per person)

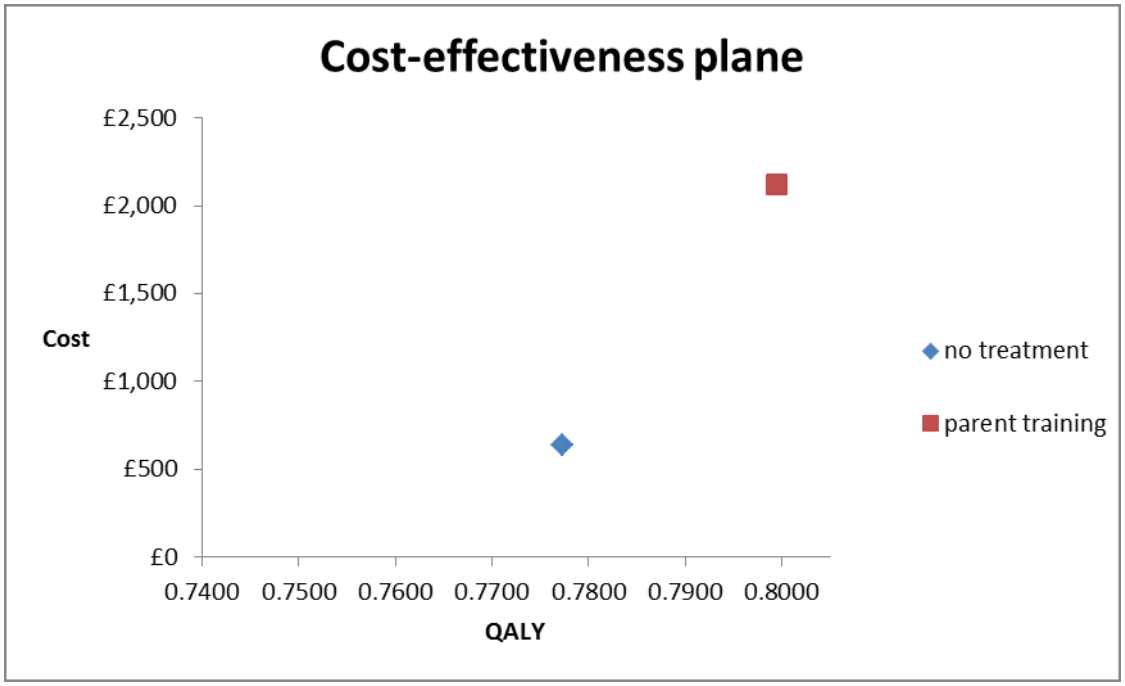
	Total cost	Total QALY
Parent training	£2,118	0.7994
No parent training	£639	0.7773
Incrementals	£1,478	0.0221
ICER	**£66,891**	

This analysis was slightly different to the others because the Pfiffner study had two timepoints and both were used in the analysis. Both the treatment effect and baseline effect fell at the later timepoint. This was the most resource intensive of the studies used in the base case analyses because it included parent and child training, and also time with the teacher, as well as family sessions and session with the family and teacher. Some of which were assumed to be 1:1 (if they were about the individual child) which adds to the costs as those costs cannot all be divided by a group. Hence although the incremental QALYs are higher than in the other analyses (because the additional response rate is higher in this study at around 30%), the larger intervention costs are causing the ICER to increase.

A threshold analysis showed that the intervention cost would have to be £606 to make the intervention cost effective.

The cost effectiveness plane can be seen below;

Figure 11: Cost effectiveness plane – base case_PFIFFNER



The PSA found that parent training has 0% probability of being cost effective.

It has been assumed in this analysis that the baseline deteriorates as well as the treatment effect. Instead it could have been assumed that the 12 week baseline effect stays at that level until the end of the model. But as only one study was used in this analysis, the effects were taken directly from the study. After the 26 week effect however then both the baseline and treatment effects are assumed to remain at that level. This raises the argument however of whether the assumptions made in the model are too strict, such as assuming the treatment effect stays at the same level. For this to be true then patients (parents/children/teachers) would have to keep applying the techniques learnt in the training, which may not be true, as there is likely to be attenuation in the fidelity with which the learned strategies are applied. Additionally, no deterioration in the baseline has been modelled (except for the period between the two timepoints in this analysis), which again may not be accurate because it is possible that over time people may get used to their medication. If treatment effect was assumed to diminish over time then this would make the treatment less cost effective. Additionally if deterioration is modelled, then relatively speaking if the treatment effect is also assumed to diminish over time then this is likely to have little an impact on the results compared to assuming they both stay at the same level.

1.3.1.5 Analysis: Base case_OSTBERG

The probabilistic results of this analysis can be found below:

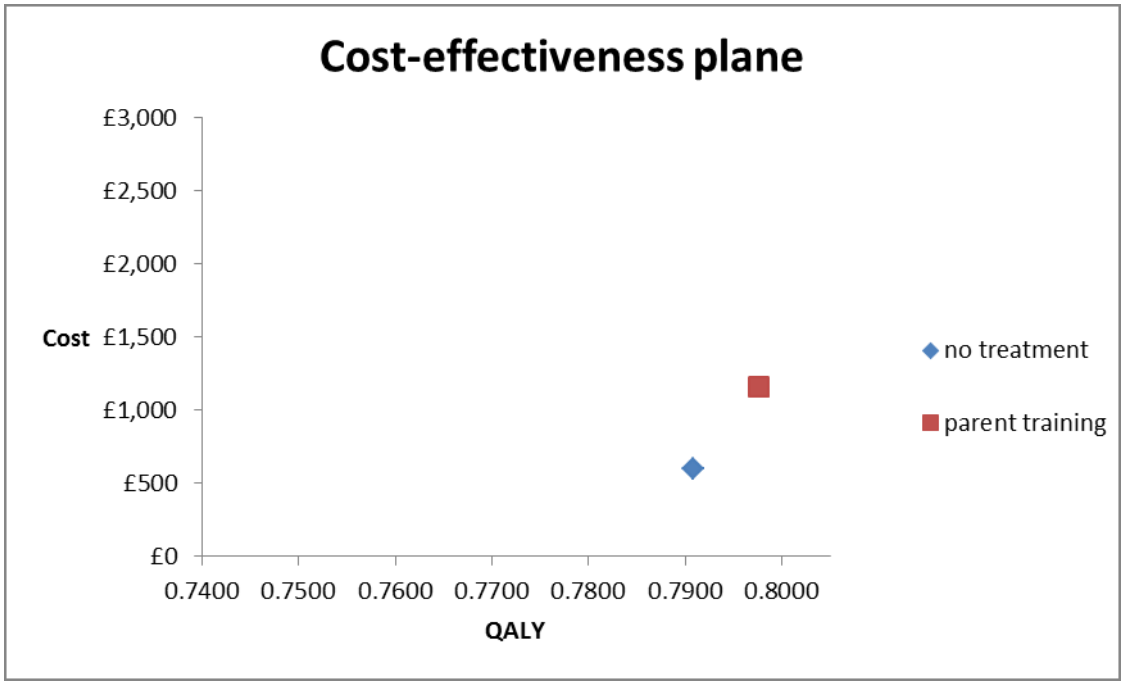
Table 13: Base case results; OSTBERG (per person)

	Total cost	Total QALY
Parent training	£1,163	0.7976
No parent training	£599	0.7908
Incrementals	£564	0.0068
ICER	**£82,915**	

The Ostberg study had a longer outcome timepoint than the other studies, at 3 months post intervention, which was around 24 weeks. This means that the linear increase in utility from non-response to response carries on from a longer period of time, so less utility is accrued because the slope of the linear line is flatter (see graphs in section 1.2.3.4). The intervention also looked at in this study was a parent and teacher intervention, which is more resource intensive than just parent training. These two factors together lead to a higher ICER than some of the other analyses. It can be seen from the cost effectiveness plane below that the total QALYs are higher than in some of the previous analyses. This is because the response probabilities from the Ostberg study were 55% (baseline) and 66%, so as they are much higher than in some of the other studies then the total QALYs accruing are larger because they are more responder. But relatively, the effects from the two arms compared to each other are quite small.

The PSA showed that parent training has a probability of being cost effective of 2% in this analysis.

Figure 12: Cost effectiveness plane – base case_OSTBERG



A threshold analysis on the intervention costs showed this would have to be below £195 to make the intervention cost effective.

1.3.2 Sensitivity analyses

1. Using the outcomes from studies that measure behavioural outcomes dichotomously, rather than symptom measures.

The results of this sensitivity analysis can be seen below.

Table 14: SA1 results

	Total cost	Total QALY
Parent training	£1,288	0.7711
Current treatment	£739	0.7601

	Total cost	Total QALY
Incrementals	£549	0.0110
ICER	**£49,944**	

Two studies have been pooled together that had behavioural outcomes because they were felt to be sufficiently similar and homogenous. The relative difference between the treatment response and baseline response is around 13% which is leading to higher incremental QALYs than some of the other analyses. As the pooled studies have been assumed to be parent and child training to be conservative, then this leads to a higher cost than just parent training alone. This is why the ICER is above £20,000.

A threshold on costs of the intervention showed that the costs of the intervention would have to be below £276 to make the intervention cost effective, keeping all other things constant.

Other sensitivity analyses that have been undertaken for the other models in the guideline include using other sources for utility values such as elicited directly. However the two alternate sources that were used in those models had either the same incremental utility or a higher incremental utility, and therefore results would either be the same as in the base case or more cost effective.

It can also be predicted what the effect would be of making assumptions about diminishing effects of the interventions after it was finished; as this would reduce the incremental QALY and therefore increase the ICER and make the interventions less cost effective.

2. Using continuous outcomes transformed into dichotomous outcomes.

Another approach that was explored in order to utilise more data on effectiveness to inform the model was the possibility of transforming continuous data to dichotomous data. There is published literature on the different methods that might be used to do this[7],[4].

Two particular methods that were being considered were; the graphical method of creating a distribution around the scale/measure that is being used to decide response, and selecting a cut off that would represent improvement or response above that cut off. The proportion of responders would then be calculated that fall into the area above the cut off. This is a method that has been employed before in NICE guidelines. The second method considered is reported in the literature and also the Cochrane handbook and is a more statistical approach which calculates the log odds ratio from the standardised mean difference. An explanation is given in Figure 13 below.

Figure 13: Conversion of SMD to odds ratio

This assumes that the continuous variable has a logistic distribution with equal standard deviations in the two intervention groups. The assumption is unlikely to hold exactly and the results must be regarded as an approximation. The log odds ratio is estimated as

$$lnOR = \frac{\pi}{\sqrt{3}} SMD$$

(or approximately 1.81×SMD).

The resulting odds ratio can then be combined with an assumed control group risk to obtain an absolute risk reduction. These assumed control group risks refer to proportions

of people who have improved by some (unspecified) amount in the continuous outcome ('responders').

This method was explored by extracting data from studies in the non-pharmacological clinical review that looked at parent training and that did not report dichotomous outcomes (but instead continuous outcomes).

The first method was ruled out because some measures do not have a validated or commonly used cut-off that we could employ. Even if there were, sometimes this is in the form of x% reduction on the scale from baseline. Such an outcome would mean that the scores of each individual would need to be known to work this out, as this isn't the same as knowing the mean of all participants before and after treatment. If individual scores were reported or if the study stated how many people met this criteria then this would already give us the dichotomous data we need. But instead it would need to be a measure where a specific cut off can be defined e.g. >6 means improvement on a scale of 1 to 10. We therefore decided to focus on the second method that would also be more statistically defensible.

After removing those studies that did not report total symptoms continuously (as we were also interested in total symptoms not subscales which most studies tend to report), we were left with two studies. These reported final values rather than change scores and had some problems;

- Both studies reported Connors ADHD index scores, one however had mean values that were much higher than the other (around 70 (before and after) whereas the other study was around 20 (before and after)). This seems very odd if they are using the same scale.
- One study (van den hoofdakker 2007) had final values that were higher in the intervention group. This is problematic because lower values mean less severe ADHD so lower values are better, and therefore although the overall change was bigger in the intervention group, as the study is reporting final values, then the interpretation that the intervention is more effective is not translating to SMD and therefore to odds ratios. Therefore although change values would be more useful, the study did not report these and calculating these from the scores before and after treatment would come with additional assumptions because this would be a post analysis calculation.

In summary; attempting to use this formulaic method of transforming continuous outcomes led to some strange results because of issues with the data as explained above, and led us to be very unconfident in the result that was being produced. After discussing this with the committee it was felt that combining this transformed odds ratio with the dichotomous data would only further increase the uncertainty in the model and further decrease the committee's confidence in the model results.

3. Using data from the under 5's population from the guideline review to assess cost effectiveness in that group.

The committee were also interested in seeing if the model could be applied to the under 5 group to see if parent training would be cost effective in that group. This could only be undertaken on the proviso that the studies included in this population group from the clinical review reported dichotomous outcomes that were needed for the model.

Four studies were included in the guideline clinical review for the under 5 age group. All four looked at some form of parent training. Two looked at the New Forest Parenting Programme

(NFPP), one looked at the Triple P parenting programme, and one looked at parent child interaction therapy.

One study did not report any dichotomous outcomes (Abikoff 2015). One study reported some dichotomous outcomes but these were on the number of children that met clinical significance points on specific subscales of the Disruptive behaviour Scale (the hyperactivity component, the ODD component), and not total ADHD symptom scores (Matos 2009). One study (Bor 2002) reports the percentage in each group that meets the reliable change index for the ECBI or the Parent Daily Report score, and that meets a 30 % reduction in observed child disruptive behaviour – both of these are behavioural outcomes rather than total ADHD symptom outcomes. The final study (Thompson 2009) reports the number of children that met a reduction in ADHD symptoms on the pre-school version of the Parental Account of Childhood Symptoms (PACS). This is the only study out of the four that reported total ADHD symptoms dichotomously. The study compared the NFPP with treatment as usual with 21 and 20 children respectively. The treatment as usual group received no treatment or they were referred onto services but they were given contact information for health visitors, GP's or school nurses which they could use as they wished. This was essentially a no treatment control group. The NFPP was an 8 week programme delivered in children's homes (8 weekly visits). The analysis on dichotomous outcomes took into account the drop outs by allocating the drop outs to non-response (i.e. intention to treat analysis). Treatment success was judged against a threshold for clinical change of a decrease in PACS ADHD symptoms of five points (the authors state this is equivalent to a 0.8 standard deviation reduction in symptoms in the current study). The outcomes reported in the study are a 40% response rate in the intervention arm (8 out of 21) and 5% response rate in the control arm (1 out of 20). It is unclear whether this relates to the outcomes measured post treatment (at 9 weeks) or those measured at a follow up timeframe (17 weeks). Also as this includes those who dropped out this is a conservative estimate. More people dropped out in the control arm than the intervention arm meaning that the control arm might be underestimating the response rate more than the intervention arm by assuming that all the drop outs would have been non responders. It is also important to bear in mind that the intervention is delivered in the children's homes and is therefore an individual intervention rather than a group intervention, which will substantially increase the costs.

Using the probabilities outlined above in the model, applying them at 9 weeks and assuming responders remain responding until the end of the time horizon (as in the base case), and also amending the costs so that the cost per person is now £1,232 (this is made up of; 8 hours of clinical psychologist time, 8 hours of prep time, 8 hours of prep time from an assistant. The cost of the assistant setting up and attending the training has been removed because it is assumed that no set up time is needed as the clinician goes to the patient's house, and also the assistant doesn't need to attend the course because it is not a group intervention) leads to an ICER of around £38,000 (deterministic analysis) making the NFPP not cost effective compared to treatment as usual.

If the intervention was a group intervention (if the cost per person from the base case analysis is assumed) then the intervention has an ICER of around £900 once the savings from the resource use avoided has been taken into account (because there is a large difference between responders and non-responders and costs would be avoided from appointments as well as the intervention being cheap because of only 1 hour per week for 8 weeks). This is however based only on one study and needs to be interpreted with caution.

1.4 Discussion

1.4.1 Summary of results

The model has shown that the cost effectiveness of parent training is uncertain. A number of studies have been used as scenarios in the base case because of the heterogeneity of studies.

When behavioural outcomes instead of total symptom outcomes were used, the ICER was also above the NICE threshold.

1.4.2 Limitations and interpretation

There a number of limitations to the model that could impact upon the interpretation;

The heterogeneity between the studies because of differences in underlying treatment status, ADHD type, intensity of intervention, and scale used to define response, meant that pooling them wouldn't be methodologically accurate because the uncertainty around the pooled result would be so large that we don't have any idea of what type of distribution that data would have, which would also cause large variability in the draws from the PSA simulations. Therefore it was decided to keep the studies separate and have various base case analyses. It is then somewhat problematic to have multiple base case ICERS as we would not be certain which one is the most accurate. However this does tell us that the ICER is very sensitive to the inputs, and that there is variation in the effectiveness of the interventions which could be down to a number of factors. Trials also tend to be fairly small which affects the quality of their interpretation as an input into the model.

The differences in the studies as touched upon above make us question their quality. The Pfiffner study for example is in mostly an inattentive subtype population of children, and had only two participants taking medication. Therefore although this study is the closest to all the studies of having a medication naïve population, the fact it is only a subtype of ADHD makes it a less applicable population, and it would be difficult to extrapolate the effect from this to the other subtypes. The studies can vary from having hardly anyone on medication (Pfiffner 2007), to having more than 70% of the study population on medication (Ostberg 2012).

As there were no drug naïve studies specifically (in a mixed subtype population), this precludes us from answering the question of whether parent training is cost effective as a first line treatment. Therefore based on the data that was available from the clinical review for the guideline, although the model is answering whether parent training (or some variation of it) is cost effective, this needs to be interpreted with caution as to how applicable these results might be to different baselines, and also whether the effectiveness of the intervention is anticipated to be the same depending on what the intervention might actually involve.

It may also be difficult to marry up the results of the model with the clinical review because the model inputs are dichotomous whereas the clinical outcomes prioritised in the clinical review are continuous. Most outcomes that involved the studies included in the model were not found to be effective in the clinical review on continuous outcomes. Some exceptions are poolings of ADHD symptoms that involve Handen 2015, Handen 2015 on the dichotomous CGI-I outcome, and the inattention outcomes from Pfiffner 2007. It may be that the different types of outcomes are capturing different aspects of the condition, or perhaps merely reflects that the arbitrary cut-offs decided for the different outcomes are not equivalent (for continuous outcomes; difference of >20% of the control group risk. and dichotomous outcomes; 50 more per 1000).

An important and related point that was discussed with the committee was whether all important effects were captured within the model (and is applicable to all models). The committee view was that the impact of behavioural therapies on the condition is not well

captured in trials. A more global function measure would be required to capture the impact on factors like self-esteem, organisation, relationships, coping with ADHD etc. and in general these more wider factors than just purely symptoms of hyperactivity and inattentiveness. Ideally quality of life or also perhaps the Clinical Global Impressions scales (CGI) are more global, but these were not as prominent in the review data as other outcomes that were more symptom based. There was therefore a strong conclusion from the committee that it is likely there are benefits from behavioural therapies that are not being captured in the model. And if in fact these were measurable and captured then this would lead to more responders which would mean more people to accrue a higher quality of life in the model.

No deterioration has been assumed for the baseline either, although it is possible that as children grow they often require dose increases based on their weight – so they can become tolerant to their prescribed dose, or may become non-adherent which could also cause deterioration, but it is possible this may be captured through the response rates for the control arms. The utilities used in the model come from a study in the Netherlands that used the UK EQ-5D tariff to elicit quality of life from parents of children with ADHD who are either responders or non-responders to medication. A limitation raised by the committee was that the quality of life of children that respond to medication may not be the same as that of children that respond to behavioural therapies, because different treatment might impact or have an effect on different symptoms associated with the condition, for example medication tends to primarily target ADHD symptoms, whereas behavioural interventions may help with symptoms but may also help with ODD symptoms and wider areas of functioning. There was however no quality of life data identified in people that were on behavioural therapy, either using relevant generic measures or applying utility elicitation methods to descriptions of health states.

It is often argued that generic quality of life measures are not sensitive enough to capture changes in quality of life associated with mental health conditions. There is no empirical evidence to suggest that the EQ-5D for example is not valid for an ADHD population.

There were a number of factors that are not included in the model that could be considered limitations. For example resource use such as how often a patient sees a GP or maybe goes to hospital if they are a responder or non-responder. Some literature has tried to investigate the cost of ADHD more generally and often compares with a group who do not have ADHD, which would not have been the right comparator to use as this would be assuming that those who did respond to ADHD consume the same resources as someone without ADHD which would probably be an underestimate. One Dutch study [22]compared the resource use of children who responded to treatment compared to those who did not respond to treatment, based on questionnaires sent out to parents to establish resource use over a one month period, and categorising the children into responders or non-responders based on descriptions given to parents. The type of resource use that was captured was that related to the condition alone and included direct medical consultations, skills training for children and parent training. The costs were extrapolated to over 12 months and found these were around £5,000 for responders and almost £8,000 for non-responders. The results of this study were not used in the model because it was felt that the resource use in a Dutch system might be more intensive than that in the UK. Also, it already includes costs associated with interventions like parent training and potentially other treatments as well, so there is probably some double counting. Some economic evaluations have included resource use associated with responders and non-responders to capture the difference. For example King 2006[14] included resource use associated with responders and non-responders that came from a survey of clinicians.

A point related to the above is how the structure of the model did not assume any further lines of treatment for non-responders, which could be seen as a limitation because in reality it is likely that if a child didn't respond to an intervention they would then try something else. There would have been many assumptions needing to be made about what the next stages of treatment might be. We know from the clinical studies that the model is based on for

treatment effect , that some of the patients were already on medication, therefore it is not a completely drug naïve group and so assuming that parent training was their first line treatment and then further assuming they follow the guideline drug pathway after that may not be appropriate. The complexities of deciding the drug pathway is quite patient specific and this is also the reason why the committee felt they couldn't model sequences of treatments in the other models in the guideline. The effectiveness of the next treatment in the sequence may also depend on how well someone responded to the previous treatment, and data on dependent probabilities is severely lacking. Additionally, omitting further treatment means the costs of any treatment people might go on to has been omitted, which may be different in the different arms because of the impact of parent training. Those on drugs will be seen at regular intervals anyway so it may be irrelevant what impact parent training has, unless it makes them stop drugs completely, or conversely – prevents them from starting drugs in the first place. There are many factors that we are uncertain of and for that reason could not be included in the model and so assumptions had to be made.

No adverse events were included in the model. It is important to note that there is a distinction between experiencing an adverse event from the intervention, and not adhering to the intervention (perhaps because the parents/children/teachers did not feel it was providing any benefit). We do not have any data on the adverse events of parent training as the adverse events review for the non-pharmacological treatments was a qualitative one. In terms of discontinuation, the studies report the average number of sessions or percentage of sessions attended, however the effectiveness should take this into account if it is an intention to treat analysis. So although we may be capturing the impact of non-attendance on outcomes, we may not be capturing the impact of this on costs by not explicitly including it in the model because the model may be underestimating the cost per family as if not all families attend then there are fewer families to spread the cost over. It is difficult to disentangle what might be direct adverse events from behavioural therapy, and what might actually be non-response and therefore trying to include adverse events in the model may results in double counting because this is already likely to be captured through intention to treat probabilities as mentioned.

In summary the model is simplistic, but adding further assumptions would only add further uncertainty.

1.4.3 Generalisability to other populations or settings

There are other similar behavioural therapy interventions that may not focus specifically on ADHD and are offered for other mental health conditions. The effectiveness of those may be different however because the purpose of the intervention might be different. Therefore there is uncertainty as to whether the results can be generalised to other populations. The intensity of the intervention can also vary and depends on the population and the setting. Whether the results can be generalisable to other countries depends on the healthcare system and costs of staff involved, as staff are the main resource involved in providing the intervention.

1.4.4 Comparisons with published studies

No published economic evaluations have been identified that looked at the cost effectiveness of parent training. This is an area severely lacking in economic evaluations.

The previous guideline model found an ICER of around £6,600 in the base case. There are a number of reasons why this is so different to the ICERs from the updated model; the QALY gains are smaller in this model because there is a smaller relative risk of response from the intervention compared to treatment. This means there is a smaller effect gain to apply a responder utility value to. The cost of the intervention is also higher in this model (and varies throughout the different base case analyses) because more components have been added

such as more preparation time and an assistant, meaning the cost per person has increased. These two effects in combination lead to an overall higher ICER in this updated model.

The NICE guideline on Antisocial behaviour and conduct disorders in children and young people: recognition and management (CG158) also conducted some original modelling looking at non-pharmacological interventions in addition to usual care compared to usual care alone. It did this separately for 3 different interventions; child focused interventions, parent focused interventions, and multi-modal interventions. The structures were markov models with time horizons of 8 or 9 years. The purpose of the analyses was to assess whether the intervention cost would be off-set by potential cost savings resulting from improvement in the behaviour of children and young people with conduct disorder. No QALYs were used. Effectiveness of the interventions were simply translated to scores on a child behavioural scale and cut-offs were used to determine if a child then moved from having conduct disorder to conduct problems or no conduct problems (the three states), with costs attached to each state. The models found that child focused treatment was cost saving, but parent focused treatment and multi-modal interventions had higher costs than the comparator. A secondary analysis on each model taking a wider perspective that also included education and criminal justice costs showed all the interventions would then be cost saving. If there is a net cost to the intervention then it is hard to compare this to the results of the models in this ADHD guideline because there are no QALYs used to derive if the additional cost is cost-effective. If costs from other public sectors were included in the ADHD guideline models it is possible that the interventions may appear more likely to be cost effective, although there are differences in the needs and behaviours of an ADHD population and a population with conduct disorder, but comorbidities are very common in mental health. Our models found that parent training cost effectiveness is uncertain, but also that the addition of behavioural therapy/CBT on to medication in people who are only partial responders (see appendix 2) are potentially unlikely to be cost effective.

1.4.5 Conclusions

The focus of this model was to evaluate the cost effectiveness of a course of parent training compared to no parent training.

The model found that parent training had a varying ICER in the base case analyses and therefore cost effectiveness is uncertain. The model is simple and has limitations such as no assumptions being made about further treatment, only being based on a few studies (and mostly single studies), which all contribute to the uncertainty in the results.

1.4.6 Implications for future research

As economic evaluations of non-pharmacological treatments is an area lacking in literature, ideally this model would encourage further research in the area.

1.5 Heterogeneity statistics for included studies

Below are the forest plots showing the study data used in the analysis, such as odds ratios used to derive the relative treatment probability, and the crude numbers of responders for the baseline arms.

Where there has been studies pooled the heterogeneity can also be seen.

Figure 14: Chacko 2009 + Handen 2015 pooled

Figure 15: Chacko 2009

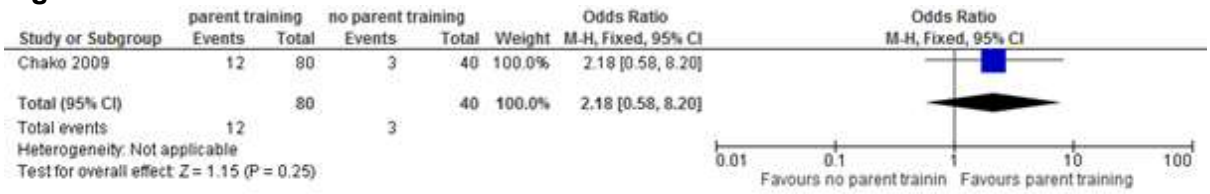


Figure 16: Handen 2015

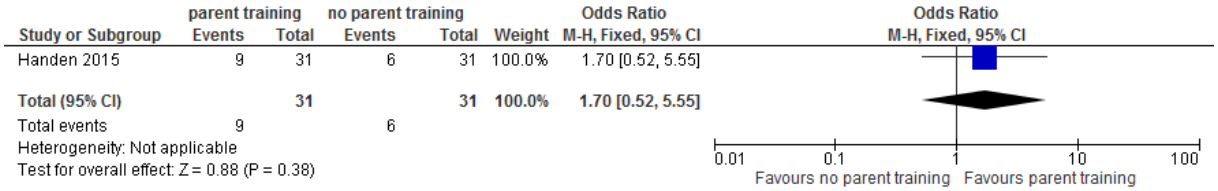


Figure 17: Pfiffner 2007 (post treatment)

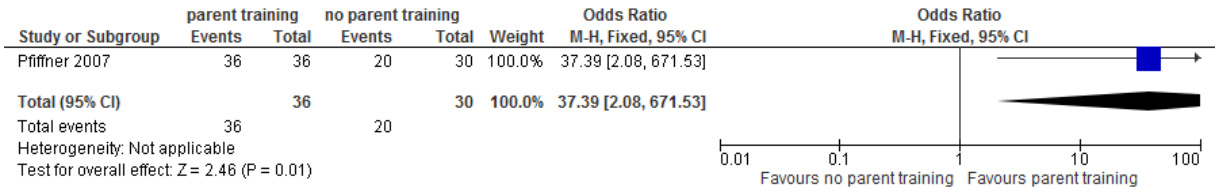


Figure 18: Pfiffner 2007 (3 months follow up post treatment)

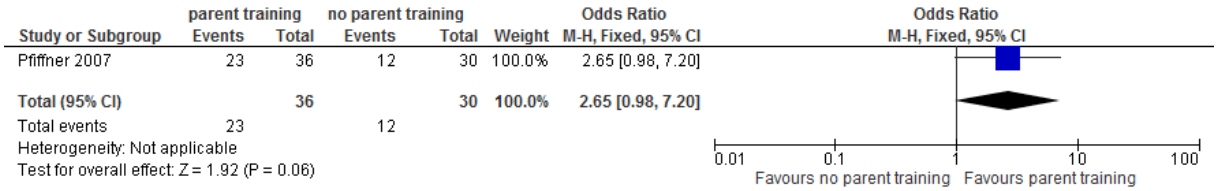


Figure 19: Ostberg 2012

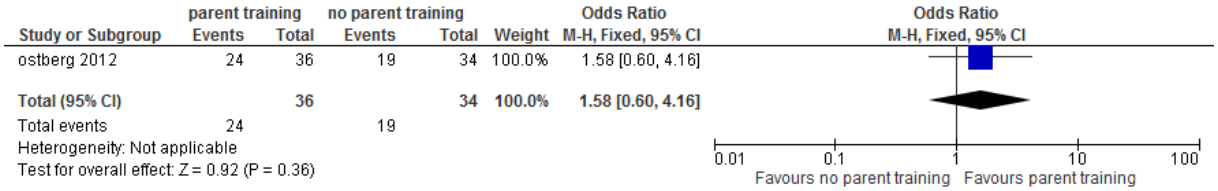
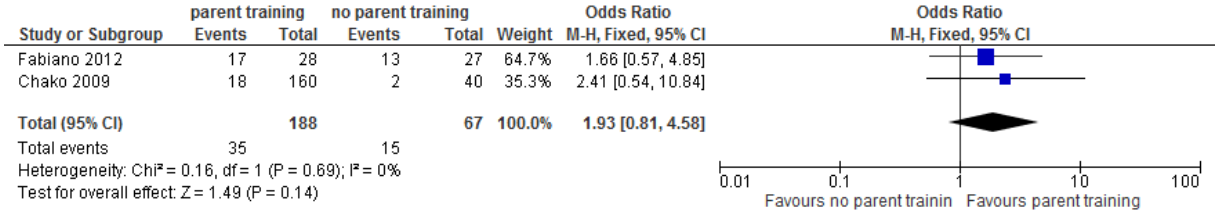


Figure 20: Behavioural outcomes sensitivity analysis



References

1. Bouwmans C, van der Kolk A, Oppe M, Schawo S, Stolk E, van Agthoven M et al. Validity and responsiveness of the EQ-5D and the KIDSCREEN-10 in children with ADHD. European Journal of Health Economics. 2014; 15(9):967-977

2. Carroll AE, Downs SM. Improving decision analyses: parent preferences (utility values) for pediatric health outcomes. Journal of Pediatrics. 2009; 155(1):21-25, 25.e21-25

3. Chacko A, Wymbs BT, Wymbs FA, Pelham WE, Swanger-Gagne MS, Girio E et al. Enhancing traditional behavioral parent training for single mothers of children with ADHD. Journal of Clinical Child and Adolescent Psychology. 2009; 38(2):206-218

4. Cochrane Handbook for Systematic Reviews of Interventions 5.1.0 [updated March 2011]. Higgins J, Green S. The Cochrane Collaboration. 2011. Available from: www.cochrane-handbook.org

5. Curtis L. Unit costs of health and social care 2013. Canterbury. Personal Social Services Research Unit University of Kent, 2013. Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2013/index.php

6. Curtis L, Burns A. Unit costs of health and social care 2016. Canterbury. Personal Social Services Research Unit University of Kent, 2016. Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2016/

7. da Costa BR, Rutjes AWS, Johnston BC, Reichenbach S, Nüesch E, Tonia T et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. International Journal of Epidemiology. 2012; 41(5):1445-1459

8. Daley D, O'Brien M. A small-scale randomized controlled trial of the self-help version of the New Forest Parent Training Programme for children with ADHD symptoms. European Child and Adolescent Psychiatry. 2013; 22(9):543-552

9. Fabiano GA, Pelham WE, Cunningham CE, Yu J, Gangloff B, Buck M et al. A waitlist-controlled trial of behavioral parent training for fathers of children with ADHD. Journal of Clinical Child and Adolescent Psychology. 2012; 41(3):337-345

10. Handen BL, Aman MG, Arnold LE, Hyman SL, Tumuluru RV, Lecavalier L et al. Atomoxetine, parent training, and their combination in children with autism spectrum disorder and attention-deficit/hyperactivity disorder. Journal of the American Academy of Child and Adolescent Psychiatry. 2015; 54(11):905-915

11. Hoath F, Sanders M. A feasibility study of enhanced group Triple P - Positive Parenting Program for parents of children with attention-deficit/hyperactivity disorder. Behavioural Change. 2002; 2002(4):191-206

12. Hodgkins P, Setyawan J, Banaschewski T, Soutullo C, Lecendreux M, Johnson M et al. Health utility scores in children and adolescents with attention-deficit/hyperactivity disorder: Response to stimulant treatment. European Child and Adolescent Psychiatry. 2013; 1):S127

13. Kandemir H, Kılıç BG, Ekinci S, Yüce M. An evaluation of the quality of life of children with ADHD and their families. Anadolu Psikiyatri Dergisi. 2014; 15(3):265-271

14. King S, Griffin S, Hodges Z, Weatherly H, Asseburg C, Richardson G et al. A systematic review and economic model of the effectiveness and cost-effectiveness of

methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents. Health Technology Assessment. 2006; 10(23):iii-iv, xiii-146

15. Lloyd A, Hodgkins P, Sasane R, Akehurst R, Sonuga-Barke EJ, Fitzgerald P et al. Estimation of utilities in attention-deficit hyperactivity disorder for economic evaluations. The Patient: Patient-Centered Outcomes Research. 2011; 4(4):247-257

16. National Institute for Health and Clinical Excellence. Social value judgements: principles for the development of NICE guidance. London. National Institute for Health and Clinical Excellence, 2008. Available from: https://www.nice.org.uk/media/default/about/what-we-do/research-and-development/social-value-judgements-principles-for-the-development-of-nice-guidance.pdf

17. Ostberg M, Rydell AM. An efficacy study of a combined parent and teacher management training programme for children with ADHD. Nordic Journal of Psychiatry. 2012; 66(2):123-130

18. Pfiffner LJ, Hinshaw SP, Owens E, Zalecki C, Kaiser NM, Villodas M et al. A two-site randomized clinical trial of integrated psychosocial treatment for ADHD-inattentive type. Journal of Consulting and Clinical Psychology. 2014; 82(6):1115-1127

19. Pfiffner LJ, Yee Mikami A, Huang-Pollock C, Easterlin B, Zalecki C, McBurnett K. A randomized, controlled trial of integrated home-school behavioral treatment for ADHD, predominantly inattentive type. Journal of the American Academy of Child and Adolescent Psychiatry. 2007; 46(8):1041-1050

20. Van Der Kolk A, Bouwmans C, Schawo S, Buitelaar J, Van Agthoven M, Hakkaart-Van Roijen L. Association of medication compliance with quality of life of children with ADHD and their parents measured by KIDSCREEN-10 and EQ-5D. Journal of Mental Health Policy and Economics. 2013; 16:S35

21. Van Der Kolk A, Bouwmans C, Schawo S, Buitelaar JB, Van Der Gaag RJ, Van agthoven M et al. The quality of life of children with ADHD and their parents: Results of EQ-5D and kidscreen-10. Value in Health. 2011; 14 (7):A294-A295

22. van der Kolk A, Bouwmans CA, Schawo SJ, Buitelaar JK, van Agthoven M, Hakkaart-van Roijen L. Association between societal costs and treatment response in children and adolescents with ADHD and their parents. A cross-sectional study in the Netherlands. Springerplus. 2015; 4:224

23. van der Kolk A, Bouwmans CAM, Schawo SJ, Buitelaar JK, van Agthoven M, Hakkaart-van Roijen L. Association between quality of life and treatment response in children with attention deficit hyperactivity disorder and their parents. Journal of Mental Health Policy and Economics. 2014; 17(3):119-129

24. Van Der Kolk A, Van Agthoven M, Bouwmans C, Buitelaar JK, Hakkaart L. Societal costs and QoL of children with ADHD and their parents: A comparison to a reference group from the general population. Value in Health. 2014; 17(7):A462

Appendices

Appendix A: Search strategy

A.1 Health Economics literature search strategy

Quality of life evidence was identified by conducting a broad search relating to ADHD population in Medline and Embase.

Table 15: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline	2008 – 28 September 2015	Exclusions Quality of life
Embase	2008 – 28 September 2015	Exclusions Quality of life

Medline (Ovid) search terms

1.	"attention deficit and disruptive behavior disorders"/ or attention deficit disorder with hyperactivity/
2.	((attenti* or disrupt*) adj3 (adolescent* or adult* or behav* or child* or class or classes or classroom* or condition* or difficult* or disorder* or learn* or people or person* or poor or problem* or process* or youngster*)).ti.
3.	((attenti* or disrupt*) adj3 disorder*).ab.
4.	(adhd or addh or ad hd or ad??hd).ti,ab.
5.	(attenti* adj3 deficit*).ti,ab.
6.	(((hyperkin* or hyper kin*) adj1 (syndrome* or disorder*)) or hkd).ti,ab.
7.	(minimal brain adj2 (dysfunct* or disorder*)).ti,ab.
8.	or/1-7
9.	limit 8 to English language
10.	letter/
11.	editorial/
12.	news/
13.	exp historical article/
14.	Anecdotes as Topic/
15.	comment/
16.	case report/
17.	(letter or comment*).ti.
18.	or/10-17
19.	randomized controlled trial/ or random*.ti,ab.
20.	18 not 19
21.	animals/ not humans/
22.	Animals, Laboratory/
23.	exp animal experiment/
24.	exp animal model/

25.	exp Rodentia/
26.	(rat or rats or mouse or mice).ti.
27.	or/20-26
28.	9 not 27
29.	quality-adjusted life years/
30.	sickness impact profile/
31.	(quality adj2 (wellbeing or well being)).ti,ab.
32.	sickness impact profile.ti,ab.
33.	disability adjusted life.ti,ab.
34.	(qal* or qtime* or qwb* or daly*).ti,ab.
35.	(euroqol* or eq5d* or eq 5*).ti,ab.
36.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
37.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
38.	(hui or hui1 or hui2 or hui3).ti,ab.
39.	(health* year* equivalent* or hye or hyes).ti,ab.
40.	discrete choice*.ti,ab.
41.	rosser.ti,ab.
42.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
43.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.

Embase (Ovid) search terms

1.	attention deficit disorder/
2.	((attenti* or disrupt*) adj3 (adolescent* or adult* or behav* or child* or class or classes or classroom* or condition* or difficult* or disorder* or learn* or people or person* or poor or problem* or process* or youngster*)).ti.
3.	((attenti* or disrupt*) adj3 disorder*).ab.
4.	(adhd or addh or ad hd or ad??hd).ti,ab.
5.	(attenti* adj3 deficit*).ti,ab.
6.	(((hyperkin* or hyper kin*) adj1 (syndrome* or disorder*)) or hkd).ti,ab.
7.	(minimal brain adj2 (dysfunct* or disorder*)).ti,ab.
8.	or/1-7
9.	limit 8 to English language
10.	letter.pt. or letter/
11.	note.pt.
12.	editorial.pt.
13.	case report/ or case study/
14.	(letter or comment*).ti.
15.	or/10-14
16.	randomized controlled trial/ or random*.ti,ab.
17.	15 not 16
18.	animal/ not human/
19.	nonhuman/
20.	exp Animal Experiment/

21.	exp Experimental Animal/
22.	animal model/
23.	exp Rodent/
24.	(rat or rats or mouse or mice).ti.
25.	or/17-24
26.	9 not 25
27.	quality adjusted life year/
28.	"quality of life index"/
29.	short form 12/ or short form 20/ or short form 36/ or short form 8/
30.	sickness impact profile/
31.	(quality adj2 (wellbeing or well being)).ti,ab.
32.	sickness impact profile.ti,ab.
33.	disability adjusted life.ti,ab.
34.	(qal* or qtime* or qwb* or daly*).ti,ab.
35.	(euroqol* or eq5d* or eq 5*).ti,ab.
36.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
37.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
38.	(hui or hui1 or hui2 or hui3).ti,ab.
39.	(health* year* equivalent* or hye or hyes).ti,ab.
40.	discrete choice*.ti,ab.
41.	rosser.ti,ab.
42.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
43.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
44.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
45.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
46.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
47.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
48.	or/27-47
49.	26 and 48