National Institute for Health and Care Excellence

Final

Attention deficit hyperactivity disorder (update)

Appendix 2: Cost-effectiveness analysis: Combination treatment in children and adolescents

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 All NICE guidance is subject to regular review and may be updated or withdrawn.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

1 Cost-effectiveness analysis: Combination treatments

1.1 Introduction

The previous model evaluating combination treatments in comparison to medication alone or behavioural therapy alone, in children, was based on two studies that directly compared the three interventions. The focus was on stimulants as the medication.

The question on combination treatments was decided as the first priority for economic modelling because there is a highly relevant trade-off with regards to whether the benefit of any additional interventions are worth the additional cost. It is also considered highly important in mental health for patients to have choices about what treatments they might prefer. Therefore, updating the previous model which sought to compare different types of treatments as well as the combination of the two, would help inform; the treatment pathway to be recommended as to whether there is a hierarchy regarding pharmacological and non-pharmacological treatments, and also whether the combination is cost effective.

There are three models replacing the previous combination model in children. The clinical data that led to three separate models is discussed in the next section below.

1.2 Clinical data overview

Ideally, to be able to fully inform what treatments are cost effective for an ADHD population (either adults or children), then all interventions would be compared to each other. This would look at interventions individually, in combination, and in sequence following non-response, taking a whole pathway approach. As such trials are not available, there is a limit to what could be compared in a health economic model. Such data would be needed because the likelihood of response from a particular treatment is believed to be influenced by what the patient may have tried before. Therefore trials that have looked at sequences or carefully selected their populations based on prior treatment are essential if the dependent probabilities are to be found. Combinations of treatments (either of different types of treatments (pharma and non-pharma), or of the same type of treatment such as different medications combined) also need to be studied in a trial because it would not be correct to assume that the effects of interventions are simply additive for example. Therefore without even at least indirect comparisons from clinical data, it would not be possible to evaluate the whole pathway in a model.

The three main questions in the guideline (non-pharmacological treatments, pharmacological treatments, and combination treatments) were all identified as economic priorities. The pharmacological review as touched on above identified many pairwise comparisons that ideally would have been combined in a Network Meta-Analysis (NMA) that would provide the effectiveness of all the treatment comparisons found. However because of issues with the populations and interventions in the studies; for example the in terms of previous treatment and response, the precise outcomes being different for different studies with a variety of scales being used. If we then wanted to try and combine data that included non-pharmacological treatments and combination into an NMA this became even more complicated still because of the precise interventions under investigation in non-pharma studies, and the level of separation between non-pharma and pharma treatments. All of this meant that the conclusions of an NMA where the indirect and direct comparisons are assumed to be assessing the same treatment effect would be difficult to interpret and would be unlikely to add significantly to the interpretation of the pairwise comparisons. There were also some published economic evaluations identified already for the pharmacological question, so because of these reasons a model on the pharmacological questions was de-

prioritised. The non-pharmacological model looked specifically at updating the previous model on parent training, as parent training was the intervention that most of the clinical data was found for, and is a commonly used form of behavioural therapy in the NHS.

This question is focusing specifically on the combination review; combination treatments versus pharma or non-pharma treatments alone, or a pharma treatment versus a non-pharma treatment. Therefore the term 'combination review' is somewhat of a misnomer because this is the review where different types of treatments were also compared with each other, as this was not a comparison that had been searched for in CG72. What becomes tricky when planning a model, is deciding what might happen to patients after they do not respond to the interventions being compared. If a drug is being compared to a combination treatment for example (for a certain period of time based on the study periods form the clinical trial), then what happens to patients who do not respond is important. In reality they may then try other treatments. This then turns into modelling sequences of treatments which as mentioned above we do not have much information on and ideally the response probabilities would be dependent. It was not possible to fully model the comparisons of sequences of drugs in the pharmacological question because of a lack of data. As a result, the pathway of drugs that a patient might follow in this guideline update has been decided based on consensus and cost considerations, therefore considering this pathway in a model is something that was already decided was not feasible for the pharmacological question and to do so in the combination question would also not be possible.

Including sequences of treatments in modelling to try and reflect reality would also lead to a model result which was not purely based on the interventions identified from the combination clinical review. A model could therefore be structured in different ways because one perspective is identifying the best 'first line' intervention is (or what should you start with given the patients placement in the treatment pathway), or another perspective might be looking at what the most cost effective intervention is as stand-alone treatments without making assumptions about what might happen next. There is a lot of uncertainty as to what treatments a clinician might get a patient to try if they do not respond to a current treatment that can be very patient specific. As touched on above this has been made via consensus for the pharmacological treatments.

Therefore there is balance between; trying to represent reality as much as possible by modelling the pathway that may occur, but at the same time keeping a model structure simple without over-simplifying the problem, and what is feasible using the data. As mentioned above – ideally we would have wanted a whole pathway model looking at starting on different treatments and testing different sequences of treatments if you are a non-responder.

What has been identified from the clinical review are various comparisons in different patient groups (children and adolescents) looking at different interventions both together (combinations) and separately. This leaves us with quite 'bitty' data. Separate models answering different questions are proposed and discussed more below. How to represent future treatments in the models then becomes quite model specific because it depends on the length of the trial data being used, what the baseline populations were in the trials and what the impact and limitations might be of including or omitting further assumptions about future treatment/resource use.

An explanation of the studies identified for this question in children as a whole will be presented below, followed by more detailed methodology of how the data was used and split up in order to answer multiple modelling questions.

As with the model on parent training, dichotomous outcomes were identified and extracted from the studies in the clinical review, as this is the only way to link to quality of life data (in other words outcome in terms of response and no response, as quality of life was only reported for responders and no responders).

23 studies were identified by the guideline clinical review looking at; Out of these 23 studies, 7 were ruled out for modelling because they were not considered to have relevant interventions for the model, as based on the non-pharma review; some interventions were felt to not be effective individually (such as neurofeedback) and if they were not clinically effective then they were not considered for cost effectiveness modelling. Out of the remaining 16 studies, 8 did not report dichotomous outcomes and were unable to be used in any modelling. This left 8 studies that had the relevant interventions and outcomes. A further 2 of these were in substance abuse adolescent populations. Substance abuse populations were listed as a subgroup within the clinical review protocols as they were felt to be a specific group whereby the results of studies in these groups may be slightly different and not gerenalisable to the general ADHD child population. It was decided with the guideline committee that these two studies should be excluded from any modelling because as there are already only a small number of studies left for modelling compared to the pool of studies identified in the clinical review, including studies that are in specific subgroups will mean the studies used for modelling are even further removed from the guideline clinical review.

The 6 remaining studies that were presented to the guideline committee and discussed for modelling can be seen in Table 1 below.

As can be seen from the table, the first 3 studies were specifically focusing on Atomoxetine as the drug being compared or combined with another treatment. So 2008 and Dose 2016 were focusing on methylphenidate as the drug being compared or combined with another treatment, with Dose 2016 being specifically in children who are stable on medication but have some remaining functional impairment. The final study, Sprich 2016, is looking at mixed drugs and is in adolescents who are stable on medication (a variety of medications) and have remaining clinically significant symptoms.

For the purposes of presenting the studies to the committee for discussion, they were split up into the following categorisations :

- Atomoxetine studies: Handen 2015, Waxmonsky 2010, Svanborg 2009
- MPH studies: So 2008, Dose 2016
- Adolescent study: Sprich 2016.

These groupings were based on previous discussions of what evidence the committee would pool from the clinical review, atomoxetine was not pooled with stimulants in the clinical review, and children and adolescents were kept separate here because CBT is seen as a different intervention to behavioural therapy.

Table 1: Studies with dichotomous outcomes

Study	Population	Intervention	Comparator	Outcomes	Notes
Handen 2015[8]	Aged 5-14. Mean age = around 8 in each group. 45.3% had received prior treatment for ADHD. The following were on melatonin for sleep; ATX+PT=7 patients, ATX= 8 patients, PT + placebo = 6 patients) Excluded people with a prior adequate trial of ATX	**Intervention 1: ATX+ parent training** Final dose of 40mg or 1.35mg/kg. Weekly 1:1 meetings of 60-90 minutes. Assumed for 10 weeks? A home visit was also conducted between the second and third session. N=31 **Intervention 2: ATX** "Final dose of 49.8mg or 1.3mg/kg. ATX doses were split twice daily to prevent side effects. Once-daily dosing was allowed if strongly preferred by a given family. ATX doses were individually adjusted according to a weight-based dosage schedule, with medical clinicians allowed to delay increases or to reduce doses due to AEs. The initial dose was 0.3mg/kg/day (rounded to the nearest 5 mg) with weekly escalations by 0.3mg/kg/day, unless there were limiting side effects or no further room for improvement, to a target dose of 1.2 mg/kg/day, and could be increased to a maximum of 1.8 mg/kg/day based on clinical status and response" N=32	**parent training (+pbo)** Weekly meetings of 60-90 minutes. 9 meetings. A home visit was also conducted between the second and third session. N=31	>=30% decrease on the SNAP and CGI-I<=2: **Intervention 1:** 0.484 **Intervention 2:** 0.469 **Comparator:** 0.290 At 10 weeks.	ITT analysis used. Also has a fourth arm of a placebo pill.
Waxmonsky 2010 [21]	Aged 6-12. Mean 8.59. Some had previously taken atomoxetine	**Intervention: ATX + Behavioural therapy** Medication provided in a single morning dose. Dose of 0.5mg/kg was started for 3 days then 0.8mg/kg for next 4 days, on day 8 everyone had their dose increased to 1.2mg/kg. At 3 weeks tolerability was	**ATX alone** Medication provided in a single morning dose. Dose of 0.5mg/kg was started for 3 days then 0.8mg/kg for next 4 days, on day 8 everyone had their dose increased to 1.2mg/kg. At	CGI-I much or very much improved. **Intervention**: 0.552 **Comparator**: 0.519	ITT analysis used.

Study	Population	Intervention	Comparator	Outcomes	Notes
	and some had started it before the trial. 37.5% had never taken stimulants. Excluded people who had previously failed to respond to atomoxetine.	assessed and dose could be increased to 1.8mg/kg if CGI-S score was 4 or worse. Mean dose at study endpoint was 1.40mg/kg in ATX + BT group. 3 components to BT; parenting program, social skills training, and school based daily report card. Sessions were weekly for 2 hrs in groups, children attended a simultaneous social skills program. N=29	3 weeks tolerability was assessed and dose could be increased to 1.8mg/kg if CGI-S score was 4 or worse. Mean dose at study endpoint was 1.47mg/kg in ATX group. N=27	At 8 weeks	
Svanborg 2009 [19]	Aged 6-15 Mean = 11.5 stimulant naïve children	**Intervention: ATX + psycho-education** 0.5mg/kg during the first week, thereafter 1.2mg/kg (< or = 70 kg) or 80 mg/day (> 70 kg). It was dispensed at 6 visits (visits 2-7) during active treatment phase. Parents participated in 4 session psycho-educational training. Four 3 hour group sessions. Could be seen as more behavioural therapy based as 'the content of the program contained core elements of more comprehensive behavioural treatment programs like parental management training (PMT) and the community parent education program (COPE)'. N=49	**placebo + psycho-education** Parents participated in 4 session psycho-educational training. Four 3 hour group sessions. Contains components that might be more behavioural training. N=50	ADHD-RS reduction of >=25% **Intervention:** 0.714 **Comparator:** 0.280 ADHD-RS reduction of >=40% Intervention: 0.633 Comparator: 0.140 Both at 10 weeks	ITT analysis used. Stated that this was an open label intervention so patients knew if they were on ATX
So 2008 [17]	Mean age around 8. All participants did not have any treatment	**Intervention: MPH +BT** 24 weekly sessions for 6 months in a group format. Three components; 1) direct management in the lab classroom. (A reward program using tokens that also included problem solving skills and anger	**MPH** Initiated at a dose of 5mg once or twice daily and increased up to a maximum of 60mg/day. Dose was increased in increments of 5-10mg. (overall	Normal ADHD symptoms on the SWAN scale (from a study of over 1000 community children scores below the 68th	ITT Sort of crossover design but outcomes from the first phase here only.

Study	Population	Intervention	Comparator	Outcomes	Notes
	history with the interventions.	management). 2) Skills training (each child training sessions lasted about 100mins. One trainer for a group of 8-9 and 2-3 assistants). 3) parent training (conducted by the author and lasted around 90 mins per session) Initiated at a dose of 5mg once or twice daily and increased up to a maximum of 60mg/day. Dose was increased in increments of 5-10mg. (overall dose range of 13.6 to 16.8mg daily). N=45	dose range of 13.6 to 16.8mg daily). N=41	percentile (0.5SD) above the mean were considered local norms). **Intervention:** 0.356 **Comparator:** 0.049 At 6 months.	
Dose 2016 [5]	Aged 6-12. On MPH on a stable dose for the previous 2 months and show functional impairment In at least one of the domains of the Weiss functional impairment scale.	**Intervention: MPH + phone assisted self help** Dose NR. N=51.	**MPH** Dose NR. N=52.	Percentage of children whose ADHD and oppositional symptom severity had shifted from a clinical to a non-clinical range (age and sex adjusted Stanin values <8) (i.e. clinically improved). On the 'symptom checklists for attention deficit/hyperactivity disorder (FBB-ADHS)' - German symptom checklist for ADHD. **Intervention**: 0.353 **Comparator:** 0.192	ITT

Study	Population	Intervention	Comparator	Outcomes	Notes
				At 12 months	
Sprich 2016 [18]	Aged 14-18, mean = 15. Adolescents Had clinically significant symptoms despite medication.	Intervention: Drugs + CBT Mean dose NR. Just states 'FDA approved medication' 12 sessions, 10 were just adolescent and therapist and 2 also involved parent. Average time for completers to complete the 12 sessions were 17/31 weeks. Two optional parent only sessions were offered as well. N=43	Drugs Mean dose NR. Just states 'FDA approved medication'. N=22	30% reduction the ADHD rating scale was used as a cut off for a treatment responder. (parent report not adolescent) **Intervention:** 0.4186 **Comparator:** 0.182 At 4 months	ITT for the CBT group (43 - includes those who crossed over from waitlist to CBT), and including all those who were originally in waitlist group (22). Outcome timeframe is unclear because it was 4 months of CBT and waitlist, and then a second phase where some waitlist people got CBT and the original CBT people got nothing, and then there was another evaluation at 8 months. But not clear for narrative dichotomous outcomes whether the outcomes were at 4 months or 8 months.

An observation that was highlighted for discussion was the outcomes in the So 2008 study. It seemed strange and unusually low that in the methylphenidate arm at 6 months only 5% of people responded. The study reported the compliance levels to the medication in each treatment arm; in the combination arm adherence to medication was 93%, and in the medication alone arm compliance was 66%. It may well be that behavioural therapy increases compliance to medication which is one of the perceived benefits, but the committee opinion was that this seemed like a substantial difference. The dose in the study was also relatively low at around 13.6 to 16.8mg. The medium dose of methylphenidate in clinical practice was reported to be 20mg per day[14] which is much lower than the dose used in western countries such as the UK. If there is a dose response relationship for pharmacological ADHD treatment, then the low dose may also go some way towards explaining the lack of response. The scale used could also have a part to play as clinical norms on the SWAN scale is not an outcome used in any other studies for comparison. After taking all of this into account, the committee felt that this study is an outlier and should be excluded from any modelling. As mentioned previously, we are already quite far removed from the main body of clinical evidence because there are so few studies left with the appropriate populations, interventions, and outcomes for modelling, and therefore it is necessary to be critical of the studies so that any models are based on studies that represent the main body of evidence as much as possible. This may mean that, if we keep categorisation of studies listed in the bullet points above as separate models, some models may have only a single study. This would not be a reason to abandon modelling however because although it may be a limitation, it is just as much a limitation to have more studies but of low quality, and pretty much all the studies identified for the guideline are small studies.

With the exclusion of the So 2008 study, this leaves 5 studies for inclusion in modelling. The committee initially felt that it would be acceptable to combine the atomoxetine children studies with the other drug children studies (not including adolescent studies). However following the exclusion of the So 2008 study, this left the Dose 2016 study that used methylphenidate in children, and the intervention was quite different to other studies because it involved families reading self-help manuals that were posted to them and telephone consultations. This study was also in people who were stable on medication but still had clinically significant symptoms, and the committee thought that it should not be combined with the atomoxetine studies because it was answering a different question. CBT should also be kept as a separate intervention and the adolescent study would also therefore inform a single study model in that population.

In summary, the modelling questions to be addressed in this review are;

1. **In children who may consider using Atomoxetine, is Atomoxetine in combination with behavioural therapy cost effective compared to Atomoxetine or behavioural therapy alone?**
2. **In children currently on MPH, is the addition of self-help behavioural therapy cost effective?**
3. **In adolescents currently on medication, is the addition of Cognitive Behavioural Therapy cost effective?**

Question 1 will be informed by the 3 Atomoxetine studies from table 1 [8,19,21].

Question 2 will be informed by Dose 2016[5]. The committee thought this to be a useful study, because although the intervention is quite different to what would be considered standard behavioural therapy in children (parents attending a course of therapy in a group or possibly individual format), it is capturing more ongoing support to families rather than a short term course, which they felt should be a minimum in practice; offering ongoing psychosocial support rather than a course of therapy followed by nothing. As it is offered on an individual basis the costs are likely to be high from the telephone consultations, but the GC thought it was worth exploring the cost effectiveness.

Question 3 will be informed by Sprich 2016. This study is in adolescents where CBT is a more age appropriate therapy. [18]

1.3 Atomoxetine combination model: Methods

1.3.1 Model overview

1.3.1.1 Comparators

Being evaluated in the model is the combination of Atomoxetine and behavioural therapy, compared to Atomoxetine alone and behavioural therapy alone. This will be informed by the three Atomoxetine studies from Table 1.

Atomoxetine dose in the model is using a maintenance dose of 1.2mg/kg per day. Behavioural therapy consists of 10 weekly sessions of 1 hour pf parent training with a clinical psychologist. Combination treatment is the sum of both these interventions.

Note that where an intervention from the studies being used has a placebo pill in combination with a behavioural therapy; for the purposes of the model this is being treated as only behavioural therapy.

1.3.1.2 Population

The population is children with ADHD, with an age range of 5-15 from the studies informing effect, with average ages of 8-11.

Further detail on the populations in the studies can be found in Table 1. In order to differentiate the population from that included in the other combination models in this document, it is important to be clear that these studies did not selectively include people only on the basis of previous response. Waxmonsky 2010 excluded those who failed ATX (meaning more likely to have included responders rather than non-responders), Handen 2015 excluded those with a previous adequate trial of atomoxetine, implying that the children in the trial are naïve to the drug in question, and Svanborg 2009 included only stimulant naïve children. Therefore contrary to the other two combination models in this write-up, these are not populations that are only non-responders, however some proprortion of each study have tried AHDH treatments before.

1.3.1.3 Time horizon, perspective, discount rates used

The time horizon of the model is 1 year, as with all the models in the guideline, because there is a lack of long term on ADHD. The trials also tend to be fairly short, and extrapolation of treatment effect over a very long time period was thought to be making too many assumptions.

Due to the time horizon no discounting will be necessary.

1.3.2 Approach to modelling

As with the parent training model, the clinical outcomes used in the model are dichotomous outcomes, as this is the only way to link to quality of life. The dichotomous outcomes are in terms of response or no response. Discontinuations also included in the model.

1.3.2.1 Model structure

The model structure is in the form of a decision tree, which can be seen below.

Because patients begin treatment when they enter they model (as that was how the trials were set up) then in the interventions that include atomoxetine, there is a probability of withdrawal from the treatment because of intolerable side effects. At the end of duration of the trials (10 weeks), patients are either classified as responders or non-responders. Responders remain on the treatment (if it involves atomoxetine, because behavioural therapy is a short term treatment) and remain responding until the end of the model.

Patients can also experience adverse events that are tolerable and do not cause them to withdraw from the treatment, but do lead to a decrement in quality of life.

If a patient withdraws because of adverse events, or does not respond to the treatment and therefore stop the treatment, then they go on to what is referred to as 'other treatment'.

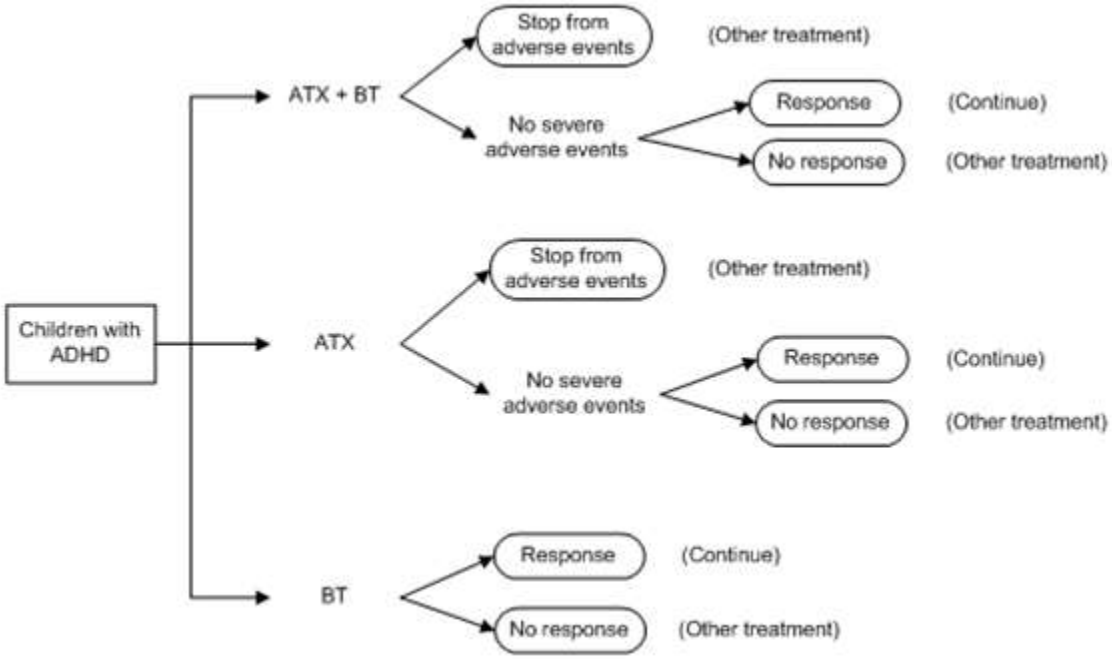
The model is limited by the fact that in reality, there are a number of treatments that people may try if they do not respond or cannot tolerate those included in this model. However there are a number of reasons why these cannot specifically be modelled;

The sequence of drugs that patients should follow in the treatment pathway until they find something that works for them is based on the pharmacological review and the economic evidence/considerations presented there. It is not the purpose of this question to decide on what sequence of treatments should be modelled. Hypothetically what would be needed to answer the question of the most cost effective pathway of treatment for the guideline as a whole would be a model comparing different sequences of treatments also involving different types of treatments (behavioural and pharmacological). The difficulty is that there is a lack of data on the dependency between treatments (i.e. the likelihood of response having tried previous treatments) which would be essential for such a model. Additionally, it was explored with the subgroup whether patients in the model should switch to the other interventions in the model if they cannot tolerate or do not respond to an intervention. The previous children combination model did incorporate switching, however it was felt that what this model is trying to answer is whether atomoxetine in combination with behavioural therapy is more cost effective than atomoxetine or behavioural therapy alone. Allowing switching in the model would have disadvantages such as; in reality patients would not necessarily go onto behavioural therapy or a combination if they failed atomoxetine, they may try another drug. Also the probabilities of response would again be independent, and the results of the model would not reflect solely the three interventions being compared.

For those reasons, it was decided that an overarching state of 'other treatment' was the best way to model this because it would not be accurate to either have people switching to the other interventions being evaluated in the model, or to begin including other interventions into the model as second line treatments (such as other drugs) for which the data is lacking. The probability of response from 'other treatment' is assumed to represent an overall probability of response in the general ADHD child population in which some people may be on a variety of treatments and some people may not be on any active treatment. The cost of 'other treatment' is represented only in terms of resource use (the number of consultations associated with responders and non-responders). This is discussed more in the resource use section, but in brief – resource use in terms of staff consultations (with a psychiatrist or nurse) is already included in the model because this is a key part of the cost of starting and continuing Atomoxetine, and therefore it made sense to continue including this resource use for the whole time horizon of the model so as not to bias against Atomoxetine or for not responding to be a cheaper outcome. The actual cost of what 'other treatment' might be such as drug costs depends on the drug and has not been included.

The structure presented below is therefore relatively simple evaluating one line of treatment for the 3 possible interventions identified from the literature of this review, for atomoxetine.

Figure 1: Model 1 structure (Atomoxetine combination model)



1.3.2.2 Uncertainty

The model was built probabilistically to take account of the uncertainty around input parameter point estimates. A probability distribution was defined for each model input parameter. When the model was run, a value for each input was randomly selected simultaneously from its respective probability distribution; mean costs and mean QALYs were calculated using these values. The model was run repeatedly – 10,000 times for the base case. Sensitivity analyses were only ran deterministically.

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 2 and in the relevant input summary tables in Section 1.3.3. Probability distributions in the analysis were parameterised using error estimates from data sources or assumptions where not available.

Table 2: Description of the type and properties of distributions used in the probabilistic sensitivity analysis

Parameter	Type of distribution	Properties of distribution
Response probabilities	Beta	Bounded between 0 and 1. As the sample size and the number of events were specified alpha and Beta values were calculated as follows: Alpha = (number of patients responding) Beta = (Number of patients) − (number of patients responding)

Parameter	Type of distribution	Properties of distribution
Adverse event/ discontinuation probabilities	Beta	Bounded between 0 and 1. Derived from the mean, and SE calculated from the confidence intervals, using the method of moments. Alpha and Beta values were calculated as follows: Alpha = mean2×[(1−mean)/SE2]−mean Beta = Alpha×[(1−mean)/mean]
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, Beta	For incremental utility of responders over non-responders: 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 For disutility associated with adverse events (as only the mean was available): Beta distribution: Bounded between 0 and 1. Derived from mean of a domain or total quality of life score and its standard error, using the method of moments. Alpha and Beta values were calculated as follows: Alpha = mean2×[(1−mean)/SE2]−mean Beta = Alpha×[(1−mean)/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 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.

1.3.3 Model inputs

1.3.3.1 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 3: 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 for behavioural therapy. Ongoing for Atomoxetine	Timeframe in the majority of clinical review studies informing treatment effect, and GC opinion.
Treatment effect	Probability of response	NMA using studies from guideline clinical review
Side effect and discontinuation probabilities	Probability of adverse events from atomoxetine Probability of discontinuation from atomoxetine	Guideline clinical review Guideline clinical review
Probability of response from 'other treatment'	Placebo response probability	Study informing treatment effect (Handen 2015)[8]

Table 4: Overview of parameters and parameter distributions used in the model

Parameter description	Point estimate	Probability distribution	Distribution parameters	Source
Response probabilities				
Probability of response from ATX+BT (at 10 weeks)	0.625		Simulations from CODA output	NMA undertaken by HE using the three treatment studies. See appendix 3.
Probability of response from ATX (at 10 weeks)	0.567		Simulations from CODA output	NMA undertaken by HE using the three treatment studies. See appendix 3.
Probability of response from BT (at 10 weeks)	0.284		Simulations from CODA output	NMA undertaken by HE using the three treatment studies. See appendix 3.
Probability of response from 'other treatment'	0.194	Beta	Alpha = 6 Beta = 25	Crude response probability from the placebo arm of Handen 2015 study.
Adverse events				
Probability of discontinuation from Atomoxetine because of adverse events	0.011	Beta	Alpha = 4.16 Beta = 373.91	Clinical review C: pharmacological efficacy and sequencing

Parameter description	Point estimate	Probability distribution	Distribution parameters	Source
Probability of adverse events from Atomoxetine	0.102	Beta	Alpha = 7.33 Beta = 64.5	Clinical review C: pharmacological efficacy and sequencing
Costs				
Atomoxetine	Cost per tablet = £1.90	NA		BNF Cost per tablet is the same regardless of dose because each pack of Atomoxetine is the same price for a variety of doses.
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 PSSRU 2016 [4]. 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) [3], 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.
Nurse (Band 7)	£130 per hour of patient contact	NA		PSSRU 2016[4] – incorporates ratio of direct to indirect patient contact.
Clinical psychologist (Band 8a)	£62 per hour			PSSRU 2016[4]
Assistant (Band 4)	£30 per hour			PSSRU 2016 [4]
Utilities				
Responder utility	0.83	Beta	Alpha = 489.7 Beta = 100.3	Van Der Kolk 2014 [20]
Non-responder utility	0.74	Beta	Alpha = 436.6 Beta = 153.4	Van Der Kolk 2014 [20]
Utility gain from responder over non-responder	0.09	Gamma	Alpha = 10.94 Beta = 0.008	Difference between responder and non-responder utility
Utility decrement for tolerable adverse events	0.02	Beta	Assuming a SE of 10% of the mean; Alpha = 98 Beta = 4801	Secnik 2005 [16]

Abbreviations: ATX = Atomoxetine, BT = Behavioural therapy

(a) Note this could also be a paediatrician. Cost per hour of a paediatrician is slightly lower than a psychiatrist so a psychiatrist cost has been used here to be conservative.

1.3.3.2 Initial cohort settings

The cohort of children begin each treatment when they enter the model. There is initially a 2 week titration period for Atomoxetine.

A small proportion of children can discontinue from Atomoxetine because of serious adverse events (in both the combination arm and the Atomoxetine alone arm). This will happen at the end of the titration period. Those who discontinue will go on to 'other treatment'. Those who have not discontinued will remain taking Atomoxetine until 10 weeks (the length of the trials effect is based on), at which point children will be classified as either responders or non-responders (applies to any intervention in the model). Responders remain responding to the treatment until the end of the model (and remain on Atomoxetine if it is one of the interventions that involve Atomoxetine). Non-responders go on to 'other treatment'.

No discontinuations are assumed to occur from behavioural therapy. Children on Atomoxetine can also experience adverse events that would not lead them to withdraw from the treatment, but they will experience a disutility associated with the adverse events.

Utility for responders will be applied linearly over the 10 week period. Consistent with the methods used in the parent training model. In the base case analysis, response is also extrapolated until 12 months. This assumption is removed in a sensitivity analysis for behavioural therapy (which is a short term treatment).

1.3.3.3 Baseline event rates

There is no baseline intervention per se in the model, as the probabilities of response used have not been applied using relative risks so as not to bias any particular intervention. Please find more detail on this in the next section. However in the network meta-analysis behavioural therapy was chosen as the baseline treatment. See appendix 3 for more information on this.

The response rate associated with 'other treatment' could be seen as a baseline. As talked about in section 1.3.2, when children stop treatment either because they discontinue or they do not respond, they go on to 'other treatment' which has been assumed to represent a general level of effectiveness in the general population who may be on a number of treatments or no treatment. This was because there would be a lack of data and it would also be too complicated to model the sequences of treatments that patients may go on to in reality as there is no defined pathway and is very dependent on the individual. The Handen study had a placebo arm where children took a placebo pill, and this has been used to represent the probability associated with this catch-all 'other treatment'. Whilst this does somewhat contradict the assumption that the placebo pills in the behavioural treatment arms of the trials will essentially be ignored, the committee felt it was a reasonable assumption to use a placebo response as the 'other treatment' response. The 'other treatment' is used to represent that children who do not respond or discontinue a treatment may then find something else that works for them, so it allows a non-responder a probability of then becoming a responder.

1.3.3.4 Treatment effects

Three studies inform the treatment effect of this model; Handen 2015, Waxmonsky 2010, and Svanborg 2009 [8,19,21].. The study details can be seen in Table 1.

Not all three studies have the same comparisons however;

Table 5: Interventions being compared in each trial

	ATX + BT	ATX	BT
Handen 2015	X	X	X
Waxmonsky 2010	X	X	
Svanborg 2009	X		X

Some other differences between the three studies are;

The populations; in Handen 2015 [8]around half had prior treatment for ADHD, in Waxmonsky 2010[21] some had taken Atomoxetine before (and the study specifically excluded those who had failed to respond to it) but some had also never taken stimulants before. Finally Svanborg 2009[19] selected stimulant naïve people specifically. These differences may have led to differences in outcomes between the studies because there is believed to be a relationship between response to a previous medication such as a stimulant and the likelihood of response to a non-stimulant.

The intensity of treatments particularly behavioural therapy; in Handen 2015 [8], a 9 week course of parent training was provided with weekly sessions on an individual basis lasting 60-90 minutes. In Waxmonsky 2010[21] the intervention had an increased intensity because there were 3 components to BT; a parenting program, social skills training for the children, and school based daily report card. Sessions were weekly for 2 hours in groups. Finally in Svanborg 2009[19], the behavioural therapy was labelled more as psycho-education, consisting of 4 group sessions of 3 hours each, although it mentions that the therapy contained behavioural therapy program components, and the committee thought that this made it more than psycho-education and is being treated here as behavioural therapy. Although there are differences between the studies in terms of behavioural therapy intensity, this model uses a behavioural therapy course of 10 weeks (of only parent training) for resource use, as in the parent training model.

Another difference between the studies is the scale being used to classify responders. The scales used are potentially important because different scales may be measuring different things e.g. more global symptoms or more behavioural aspects. Handen 2015 uses a 30% or more decrease on the SNAP scale and CGI-I rating of 1 or 2. 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 ADHD symptom focused. Waxmonsky 2010 is also using the top two tiers of the CGI-I to classify children as responders. Svanborg 2009 uses an ADHD-RS reduction. It reports two levels (>=25% and >=40%) but the 25% level has been used here because this is a commonly accepted cut-off on this scale. 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.

As touched on earlier, where a placebo pill is used in combination to a behavioural therapy in the studies being used for treatment effect, the effect is assumed to only come from the behavioural treatment. Whilst it is acknowledged that a placebo effect can be a real effect, and there are studies (this is discussed more for adults in section 1.12) that attempt to provide also sham behavioural therapies as well, it is also widely accepted that a placebo is an acceptable control arm in drug trials, and this assumption has been made for simplicity.

A network meta-analysis was undertaken to combine the data from the three trials because there was a closed loop a one study had 3 arms. Please see appendix 3 for further details on the NMA.

1.3.3.5 Adverse events

The model includes the probability of discontinuation associated with adverse events from atomoxetine (from either intervention that contains atomoxetine – combination or atomoxetine alone) (behavioural therapy is assumed to not have a probability of discontinuation). The source of this is from the clinical review using the outcome of 'discontinuation due to adverse events < 3 months' (informed by 16 studies)(see evidence report C: pharmacological efficacy and sequencing).

Although the treatment effect used is on an intention to treat basis, thereby assuming that those with no outcomes are non-responders. Because of this, it is possible that there is some double counting going on because those that go on to discontinue are actually already accounted for by being treated as non-responders (for example; in the waxmonsky study 5 people discontinued before the end of the trial in the combination arm. The intention to treat analysis took 29 participants as being the total number of people in the combination arm as 29 entered the trial in that arm. If these were excluded from the analysis (i.e. not an intention to treat analysis) then only 24 would be the denominator of the response probability. And so by including those exclusions and treating them as non-responders (because the denominator is larger) means that those 5 people are counted as both non-responders and as discontinuing).

It was felt however, that the discontinuation from adverse events due to atomoxetine should be incorporated because;

- This would make use of data from the clinical review that pooled 16 studies, which is more reliable than data from only the studies of treatment effect.
- One of the three studies used for clinical effect in the model reported no discontinuations, and the small samples in the studies informing treatment effect could also not be powered enough to detect discontinuations, therefore using the pooled clinical review data is more accurate.
- By including discontinuations, we can also capture the fact that those who do discontinue will not accrue the costs of the treatment anymore.
- Including discontinuations is more realistic in terms of model structure because an active drug is likely to be less tolerable than a behavioural therapy, and this can then be reflected in the number of people completing each of the interventions. The response probabilities are only from three studies as mentioned, and therefore although the response probabilities would have been higher if intention to treat analysis was not used (because the denominator would be smaller), we cannot be completely confident that these reflect reality anyway as there is a lot of uncertainty around them. The treatment effect probabilities used in the model will be tested through probabilistic sensitivity analysis.

Also included is the probability of tolerable adverse events. These are adverse events that would not lead to the discontinuation of treatment, but are assumed to remain for the remaining time the person is on treatment in the model. Again this is only included for atomoxetine. The tolerable adverse events are from the pharmacological safety review (evidence review D) from the outcome of 'Overall participants with adverse events < 3 months' (informed by 5 studies) Adverse events were discussed for behavioural therapy, and including these in the model was explored. However, unlike with a drug, it is more difficult to be certain that adverse events reported in a study are attributed directly to the intervention. For example it is unclear if adverse events reported in the studies for behavioural therapy that are physical (e.g. stomach ache) are simply the baseline level of adverse events that may have occurred anyway. Or if they are more behavioural adverse events, then are those in fact the symptoms of non-response or the intervention not being followed, rather than a direct adverse event? Therefore it is much harder to be confident that any adverse events studies might attribute to a behavioural therapy arm are not in fact because the intervention is ineffective. With the drugs, the trials included tend to control with a placebo comparison

and therefore one can be more confident that there are attributable adverse events above those that might be experienced at baseline. As the response probabilities being used are using intention to treat data, then the response probability is incorporating those who did not complete the trial and treating them as non-responders. Therefore, including any adverse events that may or may not be attributable to behavioural therapy could be double counting those already classified as non-responders as also having adverse events which may just be because of non-response. Attaching a quality of life decrement to those people may lead to an underestimation in their quality of life when they are already being treated as a non-responder.

Whereas with atomoxetine, because we are more certain that the adverse events are due to the drug, this probability will be included in the model explicitly so as to apply a decrement in quality of life because of adverse events specifically.

No costs will be assigned to the treatment of adverse events.

1.3.3.6 Utilities

Utilities used in the model for a responder and a non-responder are from Van Der Kolk 2014[20]. More detail on the child quality of life studies identified from a quality of life systematic search can be found in section 1.2.3.5 of the parent training model write-up (appendix 1). The study used the UK EQ-5D tariff, and was a fairly large sample.

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 is however no quality of life data associated with behavioural therapies for an ADHD population.

The utility gain associated with response is applied to all interventions in the model and also to those who respond to the 'other treatment' that patients might go on to.

As adverse events are incorporated into the model, there should also be a disutility associated with them to capture this difference in adverse events between interventions in the outcomes. No studies from the literature review on QoL (searching since the last guideline) identified quality of life associated with adverse events, but there were two studies looking at the impact of side effects on quality of life identified for the previous guideline (refs). Secnik 2005 had more participants and also elicited utilities based on stimulants or non-stimulants, so was used here. The study produced utility scores by interviewing 83 parents of children with ADHD in England. Parents were asked to value their child's current health plus 14 hypothetical health states, using the standard gamble technique.

The difference between a responder on non-stimulants with and without side effects is 0.02, and between a non-responder on non-stimulants with and without side effects is 0.01. The maximum of this will be taken and applied in the model; a utility decrement of 0.02 for tolerable side effects. This will be applied for the entire time a child remains on the treatment.

In terms of some assumptions applied in the model regarding utilities:

If a patient discontinues ATX, then no improvement in utility from the treatment is assumed and the disutility is applied for the titration period (2 weeks). For responders the utility is applied linearly over 10 weeks (the length of the trials used for effect), and then responders are assumed to remain responders for the remainder of the model. It could be argued that the response to behavioural therapy may be slower than that of a drug, and so different timeframes of response should be applied to different treatments - however Atomoxetine is a slow acting drug and it can have an onset of action within 1–2 weeks of starting treatment, but there is an incrementally increasing response for up to 24 weeks or longer[2]. So assuming a linear increase in utility gain over 10 weeks for all treatments was not felt to be a problem.

'Other treatment' also includes a utility gain because there is a response probability associated with the 'other treatment' in the model. This has been assumed to be an instant response rather than improve linearly because; it is meant to be a 'catch-all' proxy for what interventions people might go on to if they fail those in the model and therefore acts as a baseline response level, but also how quickly a response can be seen from a new treatment really depends on the treatment.

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 (see section A.1 for dates), 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.

Alternative sources of data for sensitivity analyses testing utility values are;

Lloyd 2011; This study undertook some qualitative interviews with a small group of young people (aged 11-16) and used information from this as well as information on baseline data from a clinical trial of drug treatments to come up with a range of descriptive health states; normal, borderline to moderately ill moderately to markedly ill, and severely ill. These health states were then valued by 100 members of the public using the time trade-off method to elicit the quality of life associated with the health states. Clinical trial data was then used to map from the different severity states (that were used as the health states) onto an outcome of response or no response, therefore allowing the utilities derived for the health states to be translated into utilities for a responder or non-responder. The utilities derived were 0.82 for a responder and 0.70 for a non-responder.

Hodgkins 2013; This was an abstract that used the Health Utilities Index 2 to measure utilities which is generic preference based quality of life measure similar to the EQ_5D. It has the following domains; sensation, mobility, emotion cognition, self-care, and pain. This is not a measure designed specifically for children, but parents or guardians of the children with ADHD completed the assessment. The utility was derived as part of a 7-week trial assessing stimulants, which had 196 participants. A number of cut-offs were used to categorise response, the one with the largest incremental gain has been used here of ADHD-RD ≥ 25%; responders = 0.899 and non-responders = 0.809.

Both of these sources still report the measurement of changes in quality of life from the patients (or in these cases a proxy of their parents), with the measurement of changes being from the public (either through direct valuation of health states or completing generic measures).

1.3.3.7 Resource use and costs

1.3.3.7.1 Intervention costs

Atomoxetine

The cost of Atomoxetine was found from the BNF[1].

A two week titration period was decided upon as this was deemed to be the norm based on the committee's experience. During the first 7 days a dose of 0.5mg/kg would be initiated, this would be increased to 0.8mg/kg for the next 7 days, and following this 14 day period, the dose would be increased to the maintenance dose level of 1.2mg/kg. Using the mean ages of the populations in the study; the Handen and Waxmonsky studies have a mean age of around 8 and the Svanborg study has a mean age of around 11. The Svanborg study makes up around 40% of the total populations from the three studies informing the treatment effect of the model. Therefore [(60%*8 years of age) + (40%*11 years of age)] = 9 years of age. The average weight of a 9 year old boy with ADHD according to ADHD growth charts the committee members use is 30kg. Therefore this weight was applied to the weight based doses to derive the dose in milligram form so costs could be calculated.

The table below describes the assumptions involved in calculating the doses used in the model. The first three columns show the data that was elicited from the committee. The fourth column shows the dose calculated from the committee estimates on dose and weight. As these doses are very close to some of the dose formulations that exist for atomoxetine, the tablet with the closest dose has been used for costing, as it wouldn't make sense to cost to the exact milligram otherwise whole tablets couldn't be used. So for week 1 of titration where 15mg would be taken; an 18mg tablet per day is assumed, and for the next 7 days a 25mg tablet is taken per day, and so on as can be seen from Table 6.

Each 28 tablet packet of Atomoxetine also has the same cost, regardless of the dose of each tablet. Therefore the total cost per week is the same regardless of the dose.

Table 6: Atomoxetine doses

time (days)	dose (mg/kg)	weight (kg)	total dose (mg)		Closest available tablet (Mg/ units)	Units/ pack Cost/ pack (£)	Cost/ unit (£)	no of tablets per day	Cost / day (£)	Cost/ week (£)
Titration period (2 weeks)										
7	0.5	30	15		18	28 tablets = £53.09	£1.90	1	£1.90	£13.27
7	0.8	30	24		25	28 tablets = £53.09	£1.90	1	£1.90	£13.27
Maintenance dose (after day 14)										
	1.2	30	36		40	28 tablets = £53.09	£1.90	1	£1.90	£13.27

Behavioural therapy

The cost of behavioural therapy was calculated based on resource use from the GC, which is 10 weekly sessions of 1 hour, with a group of 10 families, as that is a common format that would be used in the NHS. Although this may mean that the effectiveness being used in the model from the three trials is based on more intensive treatment than what the costs are accounting for, this is a limitation of the model because there has to be a base case level of resources, and all the studies have differences in the intensity of behavioural therapy provided.

Table 7: Behavioural therapy cost

Component	Cost (a)	Description
Clinical Psychologist		

Component	Cost (a)	Description
Set up time	£62 per hour * 10 hours = £620	1 hour for every session would be spent preparing = 10 hours of preparation (GC assumption)
Teaching time	£62 per hour * 10 hours = £620	10 sessions
Assistant		
Set up time	£30 per hour * 10 hours = £300	1 hour for every session would be spent setting up = 10 hours of set up time (GC assumption)
Admin time	£30 per hour * 10 hours = £300	The administrative tasks involved of contacting parents, inviting them and arranging them to attend the course would take the same number of hours as there are sessions being provided = 10 hours of admin time (GC assumption)
Attending course	£30 per hour * 10 hours = £300	Assistants also attend the course to help out where necessary. 10 sessions (GC assumption)
Total cost	**= £2,140**	**Total cost of providing the group parent training for 10 weeks**
Cost per family	**= £214**	

(a) Note that for some of the staff costs in the guideline (such as psychiatrist), hourly PSSRU costs have been calculated to include the ratio of direct to indirect patient time, whereas for the behavioural therapy costing the GC estimates of providing the intervention have been used with set up/prep time being elicited from the committee and costed separately. This method was used because the cost being obtained was that of a specific intervention, rather than a routine activity such a consultation, and hence the committees experience on the total time that would be involved in such an intervention was used.

Those who discontinue in the combination arm still accrue the cost of the behavioural therapy.

The cost of combination treatment will be the sum of the two monotherapy interventions. See also the next section for detail on additional staff resource use involved in the interventions.

1.3.3.7.2 Staff costs

The PSSRU 2016 was used as the source of staff costs.

The staff costs incorporated into the model are those of a hospital nurse and those of a psychiatric consultant (see **Table 4**). It may be a paediatric consultant that sees the child for a consultation rather than a psychiatrist, however the costs of the two types of consultant are very similar (£2 difference per hour) and therefore the psychiatric consultant cost has been used here because it is the slightly higher one.

Staff resource use is applied differently in the model depending on the population it is being applied to; for example someone who has discontinued Atomoxetine because of adverse events during the titration period may have used more resources than someone who hasn't discontinued during the titration period because it is expected they (or really their parents) will have had more contact with staff if the adverse events were serious enough for them to discontinue. Also, staff contact for someone responding to a drug will most likely be less than someone who hasn't responded because they will need more monitoring and support to find something that works for them and monitor the progress of any new treatments.

All resource estimates below have been elicited from committee members. The staff resources used in the model have been split below by interventions and response.

Table 8: Staff costs related to Atomoxetine and Behavioural Therapy

Resource detail	Length of contact	Staff member	Type of contact (a)	Cost
Atomoxetine				
During titration				
Initial contact (when put on the drug)	1 hour	Psychiatrist	Face to face	£208
Phone call with a nurse	20 minutes	Nurse	Telephone	£43.33
Phone call with nurse (only applies to those discontinuing from AE)	20 minutes	Nurse	Telephone	£43.33
			Total	**Discontinue = £295** **Do not discontinue = £251**
Post titration (maintenance period)				
Follow up at 5 weeks	20 minutes	Psychiatrist	Telephone	£69
Follow up at 7 weeks	1 hour	Psychiatrist	Face to face	£208
Contact with school at 7 weeks	20 minutes	Psychiatrist	Telephone	£69
Follow up at 6 months (only applies to responders)	1 hour	Psychiatrist	Face to face	£208
Follow up at 10 months (only applies to responders)	1 hour	Psychiatrist	Face to face	£208
			Total	**Responder = £763** **Non-responder = £347**
Behavioural Therapy (only applies to responders of BT alone)				
Follow up at 5 months	1 hour	Psychiatrist	Face to face	£208
Follow up at 9 months	1 hour	Psychiatrist	Face to face	£208
			Total	**= £416**

(a) Note there is no differentiation in cost between face to face contact and telephone contact, but the type of contact is listed for clarity.

Where there are no caveats in the table above as to whether it applies only to responders or non-responders, then it applies to both. The behavioural therapy consultations only apply to responders of the BT alone arm of the model because those that are on combination treatment will already have ongoing consultations if they responded to the combination treatment because they remain on atomoxetine for the remaining time in the model. Therefore there is not assumed to be any duplication of resources if a patient is on the combination treatment.

The cost per consultation for a psychiatrist is from the cost per hour in PSSRU 2016[4] (£106 per hour). However to incorporate indirect time that would be attributed to appointments (in terms of administrative time spent dictating notes, contacting other staff etc), rather than micro-costing this, a method that existed in older versions of the PSSRU of providing ratios of time spent on direct to indirect time on: patient-related activity was used (1:0.95). So each hour spent with clients requires 1.95 paid hours.

Those that do not respond to any of the treatments or discontinue drug treatment go on to 'other treatment', the resource use of which can be seen below in Table 9, depending on if they respond or do not respond to the 'other treatment'.

Table 9: Staff costs related to 'other treatment'

Resource detail	Length of contact	Staff member	Type of contact (a)	Cost
If don't respond to other treatment				
Follow up at 4 months	1 hour	Psychiatrist	Face to face	£208
Follow up at 6 months	1 hour	Psychiatrist	Face to face	£208
Follow up at 8 months	1 hour	Psychiatrist	Face to face	£208
Follow up at 10 months	1 hour	Psychiatrist	Face to face	£208
			Total	**£832**
If respond to other treatment				
Follow up at 4 months	1 hour	Psychiatrist	Face to face	£208
Follow up at 8 months	1 hour	Psychiatrist	Face to face	£208
			Total	**£416**

As discussed earlier, 'other treatment' has been used as a way to capture the difference in resource use between responders and non-responders, so as not to bias against atomoxetine which includes the resource use of consultations a part of the intervention. It also reflects reality more by assuming that a proportion of the non-responders will go on to something that they respond to in the future. It does not include the direct cost of what other treatment might be such as drug costs, as this would depend on the medication. It is assumed that for someone to be trying atomoxetine they must have already tried methylphenidate (which is usually first line in current practice), therefore to try and make assumptions about what treatment might be next in the sequence following atomoxetine would be too uncertain, and probabilities are also likely to be dependent. Therefore it is a limitation of the model that costs of the true pathway following the non-response to the interventions in the model may not have been captured accurately.

1.3.4 Computations

The model was a decision tree model constructed in Microsoft Excel 2010, and evaluated for a cohort of 1000 children.

Patients start at time zero and have the interventions for 10 weeks. For interventions that contain Atomoxetine there is an initial two week titration period as part of the 10 week intervention. A small proportion of patients can discontinue in the two week period and go on to 'other treatment'. After 10 weeks patients are assigned as being responders or non-responders based on effectiveness from trials included in the clinical review. If children respond to an intervention that contains Atomoxetine then the child continues on Atomoxetine (in the combination or Atomoxetine alone arm). Patients on Atomoxetine can also experience adverse events that do not prevent them from continuing the drug and these adverse events are assumed to remain for the remainder of the model.

Behavioural therapy has a short term timeframe because it is a course rather than ongoing like drug therapy. The proportion of people that are responders to any of the treatments are applied the responder utility linearly over the 10 weeks to represent a slowly increasing level of benefit (through utility) from that of baseline (non-response utility) to that of a responder. Responders remain responding for the remaining time in the model. Non-responders go on to 'other treatment' for the remaining time in the model.

Response probabilities were derived from a network meta-analysis of the three trials and applied in the model.

No discounting was applied because the model has a one year time horizon.

Total costs and QALYs are the sum of the costs (assumed to remain static as they are based on national sources) and QALYs in each arm. The sum of QALYs also includes the subtraction of disutility that comes from adverse events.

In the probabilistic analysis, only the QALYs are probabilistic because costs will not vary. The probabilistic cost per QALY was calculated by taking the average QALY per arm from all the simulations and finding the incremental and dividing the incremental costs by this incremental QALY.

1.3.5 Sensitivity analyses

1. Assuming the response from behavioural therapy decreases linearly from the end of treatment to end of the model for BT alone and combination arms.

This means that in the combination arm, this is applied for those additional responders over and above the number who would respond from ATX alone. For the behavioural therapy alone arm this decline in quality of life gain is applied to all the responders.

2. Behavioural therapy on an individual basis rather than a group.
3. Using alternative sources of utility data.

1.3.6 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 NGC; this included systematic checking of the model calculations.

1.3.7 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

When there are more than 2 comparators, as in this analysis, options must be ranked in order of increasing cost then options ruled out by dominance or extended dominance before calculating ICERs excluding these options. An option is said to be dominated, and ruled out, if another intervention is less costly and more effective. An option is said to be extendedly dominated if a combination of 2 other options would prove to be less costly and more effective.

It is also possible, for a particular cost-effectiveness threshold, to re-express cost-effectiveness results in terms of net monetary benefit (NMB). This is calculated by multiplying the total QALYs by the threshold cost per QALY value (for example, £20,000) and then subtracting the total costs (formula below). The decision rule then applied is that the intervention with the highest NMB is the most cost-effective option at the specified threshold. That is the option that provides the highest number of QALYs at an acceptable cost.

$Net\ Monetary\ Benefit\ (X) = (QALYs(X) \times \lambda) - Costs(X)$ *Where: λ = threshold (£20,000 per QALY gained)*	Cost effective if: • Highest net benefit

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.

1.3.8 Interpreting Results

NICE's report 'Social value judgements: principles for the development of NICE guidance'[15] 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.

As we have several interventions, we use the NMB to rank the strategies on the basis of their relative cost-effectiveness. The highest NMB identifies the optimal strategy at a willingness to pay of £20,000 per QALY gained.

1.4 Results

1.4.1 Base case

The probabilistic base case results can be seen below in Table 10, ranked in order of cost. The incremental difference (in costs and QALYs) shows the incremental cost and QALY of each intervention compared to the intervention in the row above it.

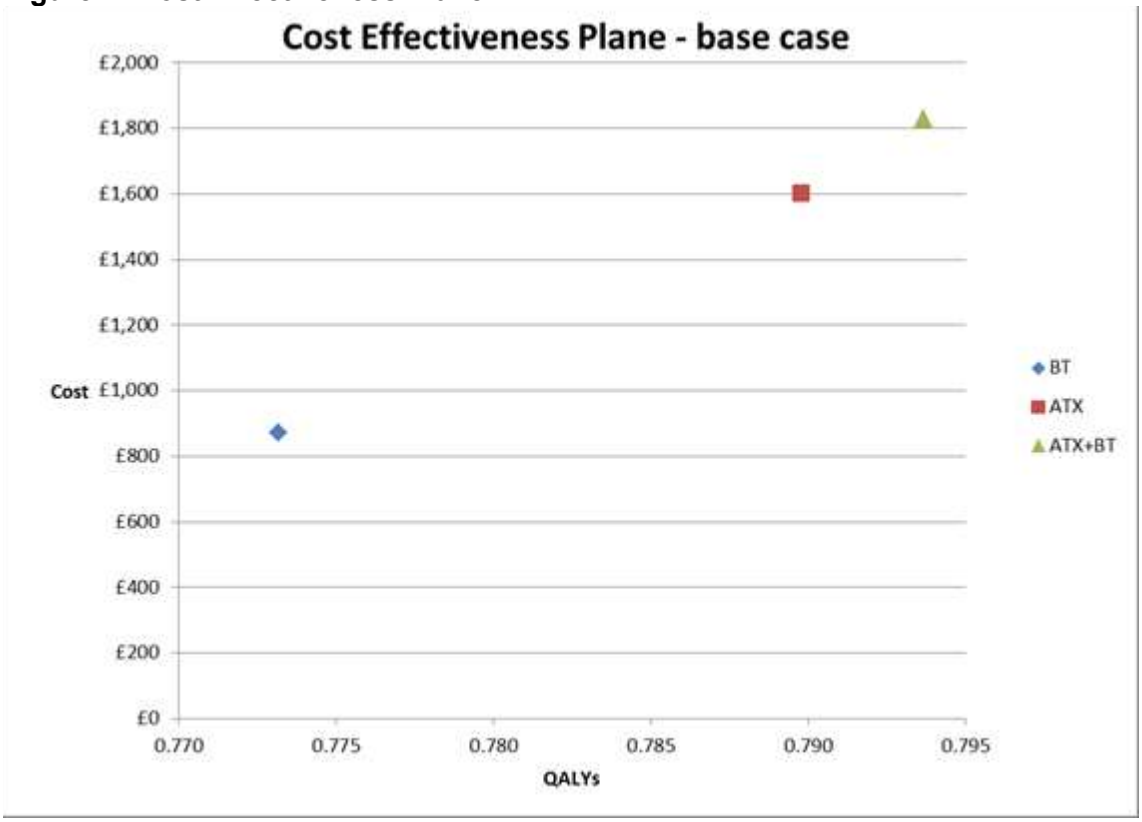
Table 10: Base case results (per person)

	Total cost	Total QALY	Incremental cost	Incremental QALY	ICER	NMB	Rank	Confidence interval of the rank
Behavioural Therapy	£870	0.773				£14,589	1	1-2
Atomoxetine	£1,602	0.790	£732	0.017	£44,175	£14,197	2	1-3
Combination treatment	£1,829	0.794	£227	0.004	£56,219	£14,051	3	2-3

The ICERS show the cost per QALY of each intervention compared to the next cheapest intervention (the row above it). Atomoxetine compared to behavioural therapy has a cost per QALY of £44,175 and combination treatment versus Atomoxetine has a cost per QALY of £56,219. The final column shows the Net Monetary Benefit (NMB) which is an alternative way of representing the intervention which is the most cost effective – that which has the highest NMB.

Figure 2 shows the cost effectiveness plane which expresses the cost effectiveness ratios graphically with the QALYs on the horizontal axis and the cost on the vertical axis.

Figure 2: Cost Effectiveness Plane



Both Atomoxetine and combination treatment have ICERs above the NICE threshold and would not be considered cost effective. We can also see from the Net Monetary Benefits that Behavioural therapy has the highest NMB meaning it is the most cost effective of the three interventions.

The simulations from the probabilistic sensitivity analysis (PSA) can also give us information about the likelihood of the interventions being cost effective. The PSA reports that behavioural therapy has a probability of being cost effective of 97% at the £20,000 threshold. This tells us the probability of behavioural therapy being the best intervention, but can also sometimes be the worst. So another method we could use is the confidence interval around the ranking of treatments, at the £20,000 threshold. If we look back at Table 10 for example then the column on the confidence interval of the rank tells us the 95% percentile of the rankings of the net benefit for all the simulations. For behavioural therapy, the confidence interval is between 1 and 2 which means behavioural therapy will always have either the first or second highest net benefit (i.e. will always be the most or second most cost effective). Looking at the confidence interval for the rank of atomoxetine, this is between 1 and 3 which means atomoxetine could be both the most cost effective and the least cost effective. Finally for the combination, the confidence interval of the rank is between 2 and 3, meaning that it will never be cost effective because it will always be the second or third most cost effective, but never the first. At the £30,000 threshold, the confidence interval for the ranking for all interventions was from 1-3, meaning that all interventions could go either be the best or the worst.

1.4.2 Sensitivity analyses

1. Assuming the response from behavioural therapy decreases linearly from the end of treatment to end of the model for BT alone and combination arms.

The results from this sensitivity can be seen below in Table 11, with treatment ordered in terms of increasing cost. These are deterministic results.

Table 11: SA1 results (per person): assuming response decreases

	Total cost	Total QALY	Incremental cost	Incremental QALY	ICER	NMB
Behavioural Therapy	£870	0.763				£14,387
Atomoxetine	£1,602	0.790	£732	0.027	**£27,197**	£14,193
Combination treatment	£1,829	0.792	£227	0.002	**£126,965**	£14,002

Behavioural therapy still has the highest NMB, although the difference is smaller compared to Atomoxetine than in the base case, and the ICER for Atomoxetine compared to BT is closer to being cost effective, showing that the results in these two different formats are agreeing with each other as they are two different ways of expressing the same result.

The reason that atomoxetine now has a lower ICER is because the total QALYs have reduced for the other two interventions but stayed the same for Atomoxetine. This is because we are saying that the effectiveness of behavioural therapy will not be maintained beyond the end of the treatment period, whereas Atomoxetine will continue being taken for the entire model period. Therefore Atomoxetine now has a higher incremental benefit to justify the additional cost over behavioural therapy alone (compared to the base case), and combination treatment now has a smaller benefit compared to Atomoxetine alone making it less cost effective than in the base case.

2. Behavioural therapy on an individual basis rather than a group.

In practice, there is a mixture of group or individual behavioural therapy provided. The previous guideline recommended group therapy on costs grounds, but recognising that there are some children that would benefit more from individual treatment for example if they are more severe. Because the outcome of the model was that behavioural therapy was the most cost effective treatment, a sensitivity analysis was conducted to see if this would change if the cost of individual therapy was used rather than group therapy.

The cost of a course of behavioural therapy for an individual was based on only the following components included in Table 7; the set up time and teaching time of a clinical psychologist, and only the administrative time of an assistant, as it was assumed that for individual sessions there does not need to be time spent setting up a room for a group of parents and children, or the assistant needing to be present to help. This led to a total cost of £1,540 per individual course of parent training as the input into the model for the cost of behavioural therapy.

Table 12: SA2 results (per person): individual behavioural therapy

	Total cost	Total QALY	Incremental cost	Incremental QALY	ICER	NMB
Atomoxetine	£1,602	0.790				£14,193
Behavioural Therapy	£2,196	0.773	£594	-0.017	dominated	£13,268
Combination treatment	£3,155	0.794	£959	0.020	£46,840	£12,718

Behavioural therapy is now dominated by Atomoxetine which is less costly and provides more QALYs.

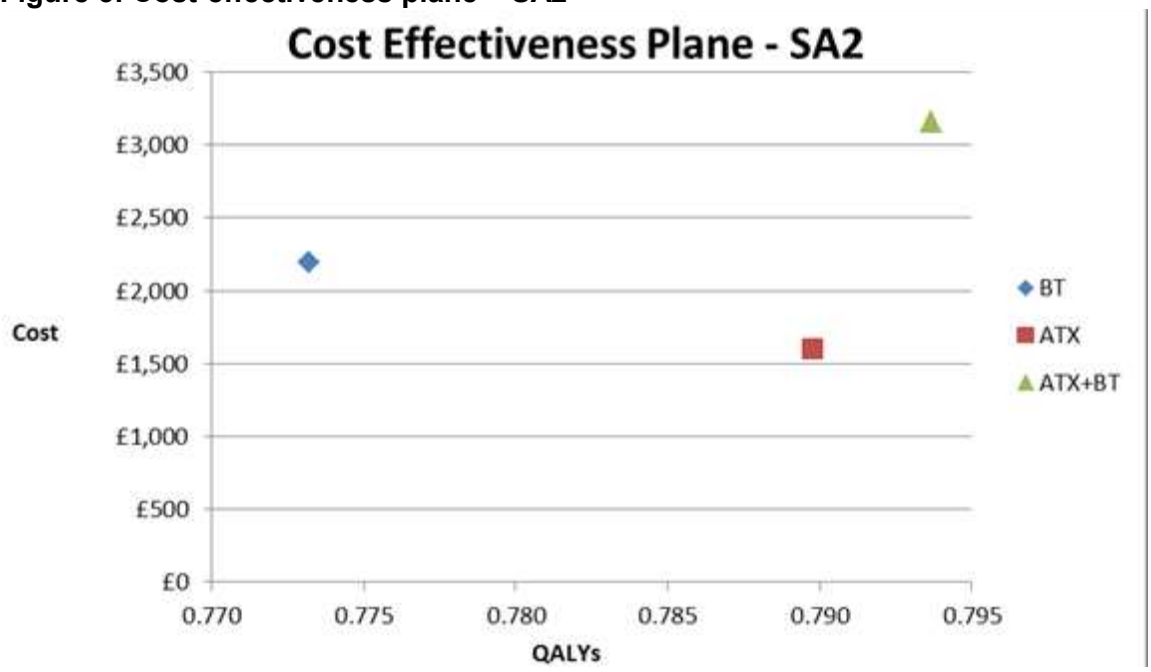
If we exclude the dominated option;

Table 13: SA2 results (per person): individual behavioural therapy - excluding dominated options

	Total cost	Total QALY	Incremental cost	Incremental QALY	ICER	NMB
Atomoxetine	£1,602	0.790				£14,193
Combination treatment	£3,155	0.794	£1,553	0.004	£399,620	£12,718

Combination treatment is not cost effective at all with a very high ICER. This is because the addition of the two treatments together is now much more expensive than the base case, so almost £400,000 will have to be spent on this intervention to gain 1 QALY, which is very cost inefficient. We can also tell by the relative values of the NMB's that Atomoxetine is the most cost effective and combination treatment is not cost effective at all. The interventions can also be seen visually on the cost effectiveness plane below because atomoxetine is to the right and lower than behavioural therapy so it creates more QALYs and has a lower cost.

Figure 3: Cost-effectiveness plane – SA2



It may be however that individual behavioural therapy is more effective than group therapy, however for behavioural therapy to be cost effective now in the model (with a much higher cost from individual therapy), it needs to have a response probability of over 80%, which is probably unlikely. In other words it needs to be more effective than both Atomoxetine and the combination using the base case response probabilities.

3. Using alternative sources of utility data

Two alternative sources of utility data were used to see if using different sources impact the model to try and alleviate the committees concerns about the EQ-5D.

Using the first alternative source of Lloyd 2011[13], the incremental QALY gain between a responder and a non-responder is 0.12. This is higher than in the base case and therefore

we would expect the cost per QALYs to be lower because an intervention with a higher response rate would now have a higher incremental QALY compared to a comparator. The results can be seen below.

Combination treatment is now closer to being cost effective than in the base case (as the ICER is lower) because the incremental QALY gain is higher as predicted.

Table 14: SA3a results (per person) using Lloyd 2011 for utilities – excluding dominated options

	Total cost	Total QALY	Incremental cost	Incremental QALY	ICER	NMB
Behavioural Therapy	£870	0.784				£14,815
Atomoxetine	£1,602	0.807	£732	0.023	**£32,122**	£14,539
Combination treatment	£1,829	0.812	£227	0.005	**£43,779**	£14,415

Using the second alternative source of Hodgkins 2013 [9], this had the same incremental gain in benefit for a responder over a non-responder as the source used for the base case utilities. Therefore this is not anticipated to have a large impact on the results. Table 15 shows the results. The ICERs are identical to the deterministic base case results because the incremental QALY has remained similar.

Table 15: SA3b results (per person) using Hodgkins 2013 for utilities – excluding dominated options

	Total cost	Total QALY	Incremental cost	Incremental QALY	ICER	NMB
Behavioural Therapy	£870	0.773				£14,594
Atomoxetine	£1,602	0.790	£732	0.017	**£44,138**	£14,193
Combination treatment	£1,829	0.794	£227	0.004	**£58,372**	£14,044

1.5 Atomoxetine combination model: Discussion

1.5.1 Summary of results

The results from the base case show that Atomoxetine and combination treatments are not providing enough additional benefit to outweigh the additional cost compared to the least effective and least costly alternative of behavioural therapy alone. This is because behavioural therapy is provided in a group format and so cost per person is lower, but also because clinical psychologists are less costly than consultants. Resource use to provide the behavioural therapy are cheaper than the drug costs and staff resource use that would be involved in providing Atomoxetine (and probably most drugs – because only when drug costs are less than a few pounds a week would the atomoxetine intervention be cost effective, as the staff costs that would be involved in checking on the patients progress and adverse events with the drug would probably outweigh the cost of behavioural therapy). It may also be that with other drugs that might be cheaper – if they are faster acting then monitoring would need to be more frequent.

As behavioural therapy is only a short term treatment, a sensitivity analysis relaxing the assumption that the probability of responding can be extrapolated to 12 months from behavioural therapy (and combination) was undertaken. This sensitivity analysis linearly reduced the QALY gain of a responder down to that of a non-responder for those patients receiving behavioural therapy from the end of the intervention to 12 months. This was done

for the cohort receiving behavioural therapy alone, and also for the additional responders to the combination over and above those responding to atomoxetine, as those on a combination would then remain on atomoxetine if they responded. This showed that if the effect of behavioural therapy (in terms of utility gain) is assumed to diminish back down to zero, then although behavioural therapy remains the most cost effective; Atomoxetine's ICER fell versus BT (because the incremental utility increased, as behavioural therapy total QALYs fell), and the ICER of combination versus Atomoxetine increased (as the incremental utility fell, as combination treatment total QALYs fell). In other words, Atomoxetine is closer to the cost effectiveness threshold (£27,197) if the benefit of behavioural therapy is not sustained.

Another sensitivity analysis was undertaken seeing how the results changed if behavioural therapy was assumed to be individual rather than group based. The cost of behavioural therapy increases significantly if the treatment is individual rather than group based because of the larger staff time commitment this would involve and not being able to spread the cost over multiple families. The results showed that behavioural therapy is now dominated by Atomoxetine which is less costly and provides more QALYs. Combination treatment is not cost effective at all with a very high ICER. This is because the addition of the two treatments together is now much more expensive than the base case. A drug treatment is now more cost effective than behavioural therapy because the costs of individual therapy are outweighing what was previously a large amount of resource use of monitoring someone on drugs.

A sensitivity analysis was also undertaken using different sources of utility data other than the EQ-5D, either using direct elicitation of utilities, or a different generic measure. These results showed that the model is sensitive to changes in quality of life, and a larger incremental gain in utility from responders over non-responders will lead to a larger incremental QALY and lower ICER.

1.5.2 Limitations and interpretation

This model aimed to compare the cost effectiveness of starting a combination of Atomoxetine and behavioural therapy, compared to starting Atomoxetine alone, or a course of behavioural therapy. Although Atomoxetine is a drug that would most likely not be at the beginning of the treatment pathway, the interventions included in the model are comparisons that were identified in the clinical review that had appropriate outcomes that could be utilised in a model, and hence the committee felt it would be a useful model. Therefore what the model is really answering is; in children who may be considering using atomoxetine, is it cost effective alone, or in combination with behavioural therapy, or is behavioural therapy alone the best choice in terms of cost effectiveness. What conclusions can be drawn from the model are highly dependent on the clinical data used, and the assumptions made about future pathways in the model and inputs such as resource use.

The clinical data informing treatment effect is based on 3 studies, with a total of 249 participants. This is a relatively small sample. The studies also have their differences in terms of the populations that they recruited, for example in one study some patients have tried Atomoxetine before and the study specifically excluded those who had not responded to it. Another study recruited stimulant naïve people specifically. These differences in populations may have had an impact on the results of the studies – for example by excluding previous non-responders to Atomoxetine, those that are in the study might be more likely to respond to the drug. The studies are also using different scales and cut-offs to categorise responders, and the intensity and sometimes content of the behavioural therapy programs is also a factor that might have caused differences in the results of the studies.

What is also important to point out is that the studies that could be used for effectiveness in all the models in the guideline are limited by the fact that dichotomous outcomes are required. This means that the data being used in the models is already potentially far removed from the overall studies identified in the clinical review and what that might be

showing as the guideline committee favoured continuous outcomes for decision making. There may be differences in what dichotomous and continuous outcomes are capturing in terms of the impact of treatment on symptoms. The cut-offs decided as minimally important differences by the committee for continuous outcomes (difference of >20% of the control group risk) and dichotomous outcomes (50 more per 1000) are arbitrary, which would be the case regardless of the level they were set. As the review could only pool pairwise comparisons; when comparing the studies the model used that look at combination treatment versus atomoxetine; dichotomous outcomes did not find combination treatment had clinical benefit (using the above criteria), but using continuous outcomes, a teacher rated outcome did find combination to be effective, and the committee prioritised teacher rated outcomes for decision making. Therefore there is some disagreement between continuous and dichotomous outcomes. When combination treatment is compared to behavioural therapy, the continuous outcomes and dichotomous outcomes agree that combination treatment is effective. Therefore there is uncertainty about whether combination treatment is more effective than atomoxetine, but more certainty that combination treatment is better than behavioural therapy.

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 particularly for behavioural therapies - the effectiveness of these on the condition are 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. The impact this may have in the atomoxetine model is to potentially increase even further the cost effectiveness of behavioural therapy, and also increase the cost effectiveness of combination treatment. It is however unclear if that would be to the level of making it cost effective versus atomoxetine.

Adverse events have been included in the model only for Atomoxetine, as it was felt that including them for behavioural therapy may be double counting if the adverse events described as behavioural were in fact because of non-response to the intervention or not adhering to the intervention. Whereas for Atomoxetine we can be more certain that adverse events would be attributed to the drug. Although the same argument could be applied to Atomoxetine that if we are using intention to treat effects then this is also likely to include those who dropped out because of adverse events, adverse events of Atomoxetine have been included because this would allow a disutility to be applied to those experiencing adverse events rather than treating them as non-responders, and also to allow them to stop the treatment and stop accruing the intervention costs. No costs have been attributed to adverse events as for the serious ones that lead to discontinuations it is assumed these will subside when the drug is stopped, and the ongoing adverse events were not separated by what type so the cost remains uncertain.

The costs of the interventions themselves could also be a limitation because the resource use estimated in the model does not necessarily align with the intensity of the interventions in the effectiveness studies, this is more so with the behavioural therapy, for example the waxmonsky study is providing parent and child training. An estimate had to be made of what typical resource use would be in the NHS, and as the effectiveness was taken from a range of studies and was rigorously tested in probabilistic sensitivity analysis, it was felt acceptable to make such assumptions. However it is possible that the cost effectiveness of behavioural therapy could be underestimated if costs should be higher to gain the effect modelled, or the effectiveness of behavioural therapy should be lower if the course provided is less intense.

A potentially large limitation is how the timeframe after the trial period is being treated in the model. As mentioned in section 1.2 at the beginning of this write-up, we would have ideally liked to model all possible interventions and sequences of treatment, but this has not proved possible. Given this, it was also difficult to decide how to structure the model in terms of the longer terms assumptions to make about what should happen to patients following failure of the interventions that we wanted to assess in the model. There were advantages and disadvantages of structuring further lines of treatment in the model. Disadvantages include; the outcomes of the model would be reflecting more than the single interventions we were trying to compare. Incorrectly capturing what the sequence might be. Using probabilities that were not dependent and therefore over or underestimating the effectiveness of further treatments. Advantages include; reflecting reality better by capturing that if patients do not respond to an intervention, they are likely to keep trying interventions until they find something they can tolerate and respond to, and therefore assuming that a non-responder remains a non-responder for the remaining time in the model may be underestimating the QALY's that will be accrued. The approach that was used in the model was to use a placebo response rate as a proxy for a general population response to a variety of treatments or no treatment, as the committee felt it would be too complicated to come up with the possible sequence of treatments, and lack of data on sequencing also made this unfeasible. This allowed the opportunity for non-responders in the model to then become responders, assuming they might try something else if they were non-responders to the interventions being compared in the model. No costs of what the treatment might be were included as this could not be defined. The costs associated with resource use were used to differentiate between a responder and a non-responder, to capture that non-responders are likely to accrue more resource use in terms of seeing staff more frequently. This method was chosen because resource use was already included for atomoxetine responders as this is part of the intervention and therefore in order to not bias against atomoxetine, the non-responders and the other interventions (BT alone) also had to have resource use incorporated. Therefore no direct costs of other treatment were incorporated, but resource use was. If responding to Atomoxetine was much more expensive compared to not responding or the other treatments then this implies it would be more cost effective to not respond to a treatment, so resource use needed to be included for all treatments and responders and non-responders. The model may still be underestimating the QALYs because we are only really including an additional line of 'other treatment' rather than having ongoing treatment changes as a possibility for non-responders. But the length of time that a person may be on a new treatment as a trial to see whether it is effective will depend on the individual treatment, and so factoring this in would also need assumptions. The cost of treatments that a patient might go on to and therefore the true cost of not responding to the interventions being evaluated in the model is also likely to be being underestimated because every time a new drug treatment is tried there will be an intense titration period where staff contact will be high (which we haven't captured as we didn't know when this might happen) and also because the actual cost of different treatments hasn't been included. Although if someone did become a responder on a new treatment they might not be having consultations as frequently as a non-responder so it could be argued that this would cost less than what is currently being estimated in the model (rather than more), however it is more likely that the cost of new treatment plus titration and monitoring costs would be more than the costs currently used for non-response. It may seem that the discussion here implies that the next treatment in a sequence would be a drug treatment, which may not be the case, but it is the most common scenario.

Overall the model structure and assumptions made about further treatment are likely to have underestimated QALYs of not responding to a treatment, and also underestimated costs. Underestimating QALYs of non-response is likely to have had more of an impact on interventions that have a higher probability of non-response i.e. BT. Underestimating costs of non-response would also have more of an impact on the interventions that have a higher non-response probability (again BT). If utilities increased but that of BT increased more than the others, and if the costs increased more for BT than they would for the other interventions, then the impact of this might be difficult to predict because higher incremental costs as well

as QALYs might imply that the other interventions would still not be cost effective compared to behavioural therapy.

The committee were also concerned about the QALY as a measure and whether this is appropriate for capturing quality of life in ADHD. There is no empirical evidence to suggest the EQ-5D is not a valid measure of health related quality of life in this area. To reassure the committee, sensitivity analyses were conducted using alternative methods and measures of capturing quality of life associated with ADHD and its changes. Differences in quality of life have an impact on the results because the incremental QALYs are very small and therefore small changes can have a large impact on the ICERs.

Only an NHS and PSS perspective was used in the model. There may be costs relevant to other public sectors that have not been included here. For example an improvement in symptoms from an intervention may mean that a child needs less educational support in school. It has also been discussed that people with ADHD are more likely to be involved in the criminal justice system (only really applies for those above the age of 10). We didn't have any information on cost savings to other sectors distinguishing between responders and non-responders. Information on criminal justice system costs avoided for example is also more long term data and long term data is lacking for ADHD in general.

Another limitation of the model is that deterioration hasn't been modelled. A patient that is responding to a treatment may in fact become tolerant to their medication because as children grow they become tolerant to their prescribed dose for example and require dose increases based on their weight. Adherence can also be a factor.

As only 3 interventions were compared in the model, and the main purpose of this question was the focus on the benefit of a combination treatment compared with those interventions in the combination separately, then the models interpretation is limited because although it tells us that behavioural therapy is the most cost effective, compared to the other interventions, it might be tentative to extrapolate this conclusion to other drugs. What also might be of interest is that if it is perhaps interpreted that behavioural therapy should be used as a first line treatment, then what might be tried next is not something this model can tell you. The previous version of this model (from CG72) incorporated switching to the other treatments in the model following non-response as part of the structure. This was discussed with the committee, but it was felt that this would not be accurate because it is not necessarily the case that if someone has failed Atomoxetine, they would then be offered behavioural therapy for example. The model has to therefore be interpreted with caution, and because the interventions compared are those that were found from the clinical review, what has been compared is limited in terms of how the actual pathway may work.

1.5.3 Generalisability to other populations or settings

Whether the results are generalisable to other treatments is unclear. For example we know from the model that even if the drug price was zero, the resource use that would be involved in terms of staff consultations and phone calls to monitor the drug still outweigh the cost of behavioural therapy. Therefore if another drug is as effective as atomoxetine, it is still likely that behavioural therapy will still be more cost effective even if it is cheaper. From the studies being used for the combination models (all 3 of them), the studies that include patients being on methylphenidate or a mix of drugs tend to have fairly low response rates (under 20%) compared to patients on atomoxetine (around 50%). Whereas the clinical review showed that a direct comparison between methylphenidate and atomoxetine on continuous outcomes did not find any difference between the two. This again potentially highlights the difference between what might be being captured on a continuous scale compared to dichotomous outcomes. However in summary, if we can assume the results could be generalisable to other drugs then a drug would have to be both cheaper and more effective than atomoxetine for it to potentially be more cost effective than behavioural treatment.

Whether the results are generalisable to other populations is also uncertain. Some other mental health conditions such as conduct disorder or depression also have treatments options available such as pharmacological treatments or psychological treatments. However as the symptoms themselves of the conditions are different, it is not possible to say whether the results of this model can be generalised.

The generalisability to other settings such as other countries would depend on how similar the health systems are. The interventions such as intensity of behavioural therapy may be different. There may also be cultural differences such as how acceptable medication is as a treatment for ADHD in some countries as the attitude towards medication for children and towards ADHD itself can also have a big impact on what treatments might be appropriate comparisons.

1.5.4 Comparisons with published studies

Literature is limited in terms of economic evaluations comparing combination treatments with individual treatments. Some economic evaluations with these comparisons were identified in the last guideline, but these have been excluded in this update because of limited applicability due to outcomes that are less relevant or the studies being out of date with regards to costs, or methodological limitations such as the studies being based on trials that have been excluded by the guideline clinical review.

The previous guideline also constructed a decision tree model comparing a combination of methylphenidate and behavioural therapy with methylphenidate alone and behavioural therapy alone. This model was structured differently to the updated model because the previous model allowed for switching; i.e. if a patient failed atomoxetine then they could switch to either behavioural therapy or the combination, and if a patient failed behavioural therapy they could again switch to either of the other two interventions in the model. The treatment effect was based on two studies that had the three comparisons relevant to the model. The results also showed that behavioural therapy was the most cost effective. The ICERS themselves are different even though the conclusion is the same because of a number of factors such as utilities that were used in the old model had a smaller incremental gain between that of a responder and non-responder. Costs were also different because of people going on to other treatments so costs are being accrued more than of just the first line interventions. Methylphenidate is also cheaper than atomoxetine.

1.5.5 Conclusions

This model was attempting to answer the question of; “In children who may consider using Atomoxetine, is Atomoxetine in combination with behavioural therapy cost effective compared to Atomoxetine or behavioural therapy alone?”

The model found that behavioural therapy is the most cost effective. This is because the other two interventions have higher costs that do not justify their benefits in relation to behavioural therapy. The model only has a simple structure and has many limitations that make the cost effectiveness of the interventions compared potentially uncertain.

1.5.6 Implications for future research

There is a lack of economic evaluations looking at combination treatments. These need to be based on trial data rather than additive probabilities of two interventions which is an incorrect assumption. More needs to be known and understood about what the impact of different treatments together on the condition is. Trials needed to inform any treatment effect in a model also need to be very clear about the populations being included in their studies i.e. whether these are drug naïve, whether they are responders, non-responders or partial responders, as it tends to be a mix. Ideally this model will inspire further research looking at the cost effectiveness of combination treatments.

1.6 MPH + self-help behavioural therapy model: Methods

1.6.1 Model overview

1.6.1.1 Comparators

This model is comparing staying on Methylphendiate (MPH) if you are a partial responder versus adding telephone assisted self-help behavioural therapy (BT). The model is therefore interested in the added value of a behavioural therapy on top of medication. The intervention involved parents reading 8 self-help booklets dealing with disruptive behaviour disorders and parenting that were mailed to them approximately every 2 weeks. Parents received 10 phone consultations of about 30 minutes each in the first 6 months, and then 4 booster calls during the second 6 months.

This is based on a single study reporting outcomes at 12 months. The GC thought that analysing the cost effectiveness of this study would be useful because it is an intervention they envisaged could be used as a baseline intervention in current practice because; it is more longer term than the usual courses of behavioural therapy, it involves a self-help and telephone consultations. Although as the intervention will be provided on an individual basis, so the cost of the behavioural therapy is likely to be high.

1.6.1.2 Population

The population is children with ADHD who are on a stable dose of MPH, but had functional impairment (in the study this was functional impairment in at least one of the domains of the Weiss Functional Impairment Rating Scale).

This can be seen as the baseline population because children are on MPH in both the intervention and the control group.

1.6.1.3 Time horizon, perspective, discount rates used

The time horizon of the model is 12 months. This also happens to be the length of the trial, however 12 months has been used in all the models in this guideline because of a lack of long term data.

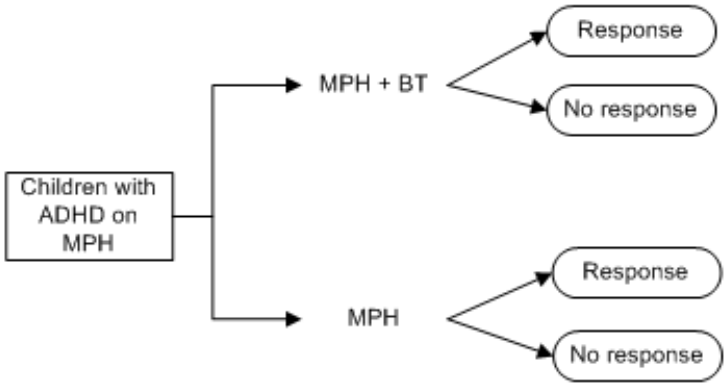
No discounting is necessary because of the time horizon of the model.

1.6.2 Approach to modelling

As with the other models, the clinical outcomes used in the model are dichotomous outcomes, as this is the only way to link to quality of life. The dichotomous outcomes are in terms of response or no response.

1.6.2.1 Model structure

The model is a decision tree which can be seen below.



Children enter the model being stable on methylphenidate, and can either remain on methylphenidate or add behavioural therapy. As the model is using a time horizon of 12 months and the trial data is also 12 months long – no assumptions need to be made beyond 12 months about what patients might then go on to.

Everyone in the baseline arm of the model stays on the baseline for the whole time period regardless of whether they respond or not, as the baseline can be seen as current practice. This was felt to be acceptable because;

- the trial period is as long as the model time horizon;
- It is reasonable that; if it was already acceptable for children to stay on a treatment that was only partially effective (and children had to be stable on MPH for 2 months prior to entry into the trial), then it must also be acceptable that children can stay on this for a longer period of time i.e. the whole time horizon of the model – which is what happens in the trial. There is no perfect response in practice and so partial response can also be considered as treatment being successful. The comparator arm is therefore being treated as a baseline, because it was meant to capture 'routine clinical care', whereas the intervention is looking at a behavioural program in addition to routine clinical care.
- The study the model is based on allowed people to change drugs in the trial and use additional treatments like support groups. Therefore from an effectiveness perspective it is assumed that trying other treatments is already included in the response probabilities. The costs however have not been captured. The danger of omitting this is that the costs are not adequately captured. It might be reasonable to assume that the underlying resource use associated with the medication is likely to be the same in both arms if patients stay on the partially effective drug for the whole time period and therefore cancel out. But there may be a difference in which drugs people changed to in the two arms or whether the behavioural therapy itself affected drug changes which wouldn't be captured as we are uncertain about this. The study implies that the intervention group ended up having fewer people on MPH and more people on other drugs – meaning that treatments people changed to are likely to be more expensive in the intervention arm (paraphrasing from the study; "*chi-square tests indicated that a significantly lower percentage of intervention group children still received methylphenidate at postassessment compared to control group children (6 cases in the intervention group and 11 cases in the control group), whereas significantly more intervention group children received lisdexamfetamine"*. It is unclear why the addition of BT has this impact, but not including this in the model means that we may be underestimating the ICER of the combination treatment. From the utility side we are also probably underestimating the utilities because people are likely to find something effective eventually if they change treatments. However including other treatments would then muddy how the ICER is interpreted because it is not capturing the costs and QALYs of solely the interventions we set out to investigate. Because of the

uncertainty around what treatments people go on to, no other treatment costs have been included here so as not to bias particular arms with assumptions that may be incorrect.

As what is being investigated here is the addition of behavioural therapy to a population that are all on medication, and the trial period was long; it was not felt to be a particularly large limitation to not make assumptions about people changing treatments.

1.6.2.2 Uncertainty

The model was built probabilistically to take account of the uncertainty around input parameter point estimates. A probability distribution was defined for each model input parameter. When the model was run, a value for each input was randomly selected simultaneously from its respective probability distribution; mean costs and mean QALYs were calculated using these values. The model was run repeatedly – 10,000 times for the base case. Sensitivity analyses were only ran deterministically.

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 16 and in the relevant input summary tables in 1.6.3. Probability distributions in the analysis were parameterised using error estimates from data sources or assumptions where not available.

Table 16: Description of the type and properties of distributions used in the probabilistic sensitivity analysis

Parameter	Type of distribution	Properties of distribution
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	*For incremental utility of responders over non-responders:* 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 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.

1.6.3 Model inputs

1.6.3.1 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 17 below. More details about sources, calculations and rationale for selection can be found in the sections following this summary table.

Table 17: Summary of base-case model inputs

Input	Data	Source
Population	Children with ADHD on a stable dose of MPH with some remaining functional impairment (age 6-12)	Study informing treatment effect
Time horizon	12 months	
Length of treatment	12 months	Study informing treatment effect
Treatment effect	Probability of response	Guideline clinical review

Table 18: Overview of parameters and parameter distributions used in the model

Parameter description	Point estimate	Probability distribution	Distribution parameters	Source
Base case effect (a)				
MPH response rate (baseline)	0.192		Simulations from CODA output	Control group response probability from Dose 2016 [5]
Combination response rate	0.361		Simulations from CODA output	From analysis of the single study (Dose 2016 [5]) using Winbugs software.
SA1 effect (b)				
SA1: MPH response rate	0.31 (16/52)			Crude control group response probability from Dose 2016, using the Reliable Change Index
SA1: Relative risk of response from intervention (adding behavioural therapy = at 12 months)	1.34			Calculated from response probability of intervention group compared to control group in Dose 2016, using the Reliable Change Index.
Cost (£)				
Clinical psychologist (Band 8a)	£62 per hour			PSSRU 2016 [4]

Parameter description	Point estimate	Probability distribution	Distribution parameters	Source
Utilities				
Base case effect (a)				
Responder utility	0.83	Beta	Alpha = 489.7 Beta = 100.3	Van Der Kolk 2014 [20]
Non-responder utility	0.74	Beta	Alpha = 436.6 Beta = 153.4	Van Der Kolk 2014 [20]
Utility gain from responder over non-responder	0.09	Gamma	Alpha = 10.94 Beta = 0.008	Difference between responder and non-responder utility

Abbreviations: PSSRU: personal social services resource unit

(a) Note that the base case baseline and treatment effect were derived by iinputting the raw values from the study into Winbugs software. The probabilistic effects were derived using CODA output from winbugs of 60,000 simulations.
(b) The sensitivity analysis 1 baseline and treatment effects were derived directly from the study (with the relative risk calculated in Revman software).
(c) The reliable change index is a statistical measure to see if the change in scores before and after treatment is statistically significant.

1.6.3.2 Initial cohort settings

The addition of behavioural therapy to MPH is compared to MPH alone. The intervention is a 12 month intervention provided individually.

The quality of life of responders is assumed to increase linearly over the 12 months 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.

1.6.3.3 Baseline event rates

The baseline event rate in the model is the response probability in the MPH alone arm. This is seen as the baseline because children who enter the model are already on this treatment and the model is therefore investigating the additional effectiveness of adding behavioural therapy and the trade-off with the additional costs.

1.6.3.4 Relative treatment effects

The treatment effect is based on a single study (Dose 2016) as this was the only study identified comparing MPH alone with a combination in the combination question clinical review for children. As mentioned in section A.2, one other study including MPH was also identified but was excluded from the model because of concerns about its population.

There were a total of 103 children in the trial which is not a very large population.

The outcome used to define response at 12 months was using the percentage of children whose ADHD had shifted from a clinical to a non-clinical range (age and sex adjusted Stanine values <8) on a German symptom checklist for ADHD (FBB-ADHS). This was regarded as showing ‘clinical improvement’. The FBB-ADHS consists of 20 items and is consistent with the DSM-IV version.

The study actually used two methods of defining response; one was that mentioned above, and the second was the Reliable Change Index (RCI). This is not a scale but a statistical measure to see if the change in scores before and after treatment is statistically significant.

The FBB-ADHS was used in the base case because it is an ADHD scale. With the results using the RCI being tested in a sensitivity analysis.

Both the intervention and comparator probability of response was derived from inputting the raw values from the study into Winbugs14 software. This method was undertaken so that the probability of response for the interventions being compared could be derived probabilistically using CODA output from Winbugs, which would keep the correlation between the baseline and treatment effect when draws were being taken from a simulation for the PSA. The code for the baseline and pairwise models from winbugs can be found in section 1.13.1. Fixed effects models were used because it was only one study.

1.6.3.5 Utilities

Utilities used in the model for a responder and a non-responder are from Van Der Kolk 2014 (ref). More detail on the child quality of life studies identified from a quality of life systematic search can be found in section 1.2.3.5. of the parent training model write-up. The study used the UK EQ-5D tariff, and was a fairly large sample.

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 is however no quality of life data associated with behavioural therapies for an ADHD population.

As with the previous model, sensitivity analyses were conducted using two alternative sources of utilities to see the impact this would have (see section 1.3.3.6 for an explanation of these studies).

1.6.3.6 Resource use and costs

1.6.3.6.1 Resource use of providing intervention

Below is the cost of the intervention of which the only component is staff time spent talking with the parents. There would also be the cost of providing the self-help booklets to parents but this was considered to be a negligible and one off cost and was not included.

Table 19: Intervention cost

Component	Cost	Description
Clinical Psychologist		
Set up time	£31 per 30 minutes * 14 telephone consultations = £434	30 minutes of prep time for every 30 minute phone call
Patient contact time	£31 per 30 minutes * 14 telephone consultations = £434	14 phone calls of 30 minutes each
Cost per family	**= £868**	

A clinical psychologist would provide the intervention as it is a behavioural therapy.

No other resource use such as consultations associated with response or non-response have been included in the model. Because the underlying population are all on medication, then they would all be monitored by a clinician (psychiatrist or paediatrician) at regular intervals anyway, then there would be no duplication of staff resources because of response or non-response to the behavioural therapy. It is not known however what impact the intervention might have on the underlying resource use which could lead to differences in costs between the two arms. in theory because this is a population already all on drugs, then unless the behavioural therapy helped them to not need drugs anymore, existing resource

use is unlikely to be affected. It may also be that the behavioural therapy leads to better compliance to medication that then makes it more effective and then perhaps responders need less frequent consultations, and would also perhaps not be considering changing treatment which would include titration and more monitoring. However all these theories would require a lot of assumptions as to what treatments people might stop, or change on to and when, or become more adherent to, in order to include these costs.

Everyone in the baseline arm of the model stays on the baseline for the whole time period regardless of whether they respond or not. So no further assumptions are made about non-responders. Please see section 1.6.2.1 for more of a discussion on this.

1.6.4 Computations

The model was a decision tree model 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.

Patients start at time zero and have the interventions for 12 months. At 12 months patients are assigned as being responders or non-responders based on effectiveness from the trial. The proportion of people that are responders to any of the treatments are applied the responder utility linearly over the 12 months to represent a slowly increasing level of benefit (through utility) from that of baseline (non-response utility) to that of a responder. Responders remain responding for the remaining time in the model. As the intervention period from the trial is the same length as the time horizon of the model, no assumptions are made about what happens to non-responders following the intervention. No assumptions are made about changes to the baseline treatment either as this is common to both arms.

Response probabilities were derived from inputting the raw numbers of respnders and total people in each arm into winbugs to derive probabilities of response (as well as uncertainty which was used to derive 60,000 simulations of the response probabilities from CODA output that were then used in a PSA).

No discounting was applied because the model has a one year time horizon.

Total costs and QALYs are the sum of the costs (assumed to remain static as they are based on national sources) and QALYs in each arm.

In the probabilistic analysis, only the QALYs are probabilistic because costs will not vary. The probabilistic cost per QALY was calculated by taking the average QALY per arm from all the simulations and finding the incremental and dividing the incremental costs by this incremental QALY.

1.6.5 Sensitivity analyses

1. Another way that the study classified response was using the Reliable Change Index. This sensitivity analyses uses those response rates.
2. Assuming effect increases linearly to 6 months as the phone calls are more intense up until that point, and stays at that level until 12 months.
3. 2-way sensitivity analysis of baseline effect and treatment effect.
4. 2-way sensitivity analysis of time horizon and utility gain
5. Using alternative sources of utility data.

1.6.6 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 NGC; this included systematic checking of many of the model calculations.

1.6.7 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

When there are more than 2 comparators, as in this analysis, options must be ranked in order of increasing cost then options ruled out by dominance or extended dominance before calculating ICERs excluding these options. An option is said to be dominated, and ruled out, if another intervention is less costly and more effective. An option is said to be extendedly dominated if a combination of 2 other options would prove to be less costly and more effective.

1.6.8 Interpreting Results

NICE's report 'Social value judgements: principles for the development of NICE guidance'[15] 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.

1.7 MPH + self-help behavioural therapy model: Results

1.7.1 Base case

Probabilistic results for the base case can be seen below.

Table 20: Base case results (per person)

	Total cost	Total QALY
MPH	£0	0.7573
MPH + BT	£868	0.7648
	£868	0.0076

	Total cost	Total QALY
ICER	**£114,803**	

The combination treatment is highly cost-ineffective, because the incremental effect cannot justify the incremental cost. The probabilistic results also showed that the probability of MPH alone being cost effective at any willingness to pay threshold was always 1 up to £24,000, and very close to 1 from £24,000 to £30,000.

Threshold analyses:

A threshold analysis on costs showed that the cost of the intervention would have to be £151 to make the intervention cost effective. This would only equate to the cost of between 2 to 3 sessions of 30 minute phone calls.

A threshold anlaysis on incremental QALYs also showed this would have to increase from 0.0076 in the base case to 0.0434 to make the intervention cost effective. The time horizon was also tested and the effect would have to be extrapolated to around 3 years to make the intervention cost effective (an assumption made there was that the effect would increase linearly until 12 months (end of treatment) and then remain at that level).

A 2-way sensitivity analysis was undertaken varying both the time horizon and the utilty gain (of responders over non-responders), results are reported under sensitivity analysis 4 in the next section.Threshold analysis were attempted on the treatment effect, but the results were showing that even with a baseline effect of nearly zero (all other things staying the same) the ICER would still be high. Also conducting a threshold analysis on the treatment effect (and keeping the baseline the same) gave an outcome that meant the response probability to the combination would be more than 1, which would not be possible as we are dealing with probabilities that must be bounded by 0 and 1. Because of these results, a two way sensitivity analysis was undertaken varying both the baseline and treatment effects to see the impact on the ICER. See the next section for the results.

1.7.2 Sensitivity analyses

1. Using the Reliable Change Index as the method the study uses to decide response.

The results of this deterministic analysis are shown below

Table 21: SA1 results (per person): using reliable change index

	Total cost	Total QALY
MPH	£0	0.7677
MPH + BT	£868	0.7724
	£868	0.0047
ICER	**£184,379**	

The ICER has fallen in this sensitivity analysis because the incremental effect is smaller.

2. Assuming effect increases linearly to 6 months as the phone calls are more intense up until that point, and stays at that level until 12 months.

Table 22: SA2 results (per person): effect peaks at 6 months

	Total cost	Total QALY
MPH	£0	0.7573

	Total cost	Total QALY
MPH + BT	£868	0.7687
	£868	0.0109
ICER	**£76,407**	

The ICER has fallen compared to the base case because the QALY gain has increased. Applying the utility gain linearly up until 6 months and then assuming the gain remains until 12 months is creating a larger area for utility gain graphically because the slope of the gain is now higher if being applied to a shorter period, as well as there now being a non-linear utility gain from 6 to 12 months.

These results show that the ICER is very sensitive to changes in the QALY gain because they are so small.

3. 2-way sensitivity analysis of baseline effect and treatment effect.

The baseline and intervention effects that were fed into the model for the two-way analysis can be seen below in Table 23. In the first column are the baseline effects used, and in the first row are the intervention effects used. In the main body of the table are the ICERs that would be derived from applying the respective baseline effect and treatment effect from that row and column. The cells that contain 'NA' mean that the intervention would be the same or less effective than that baseline.

Table 23 also shows the results of the sensitivity analysis. There are no ICERs that are below £20,000. The closest is when the baseline response probability is 0.1 and the intervention response probability of 0.95.

Table 23: 2-way sensitivity analysis – ICERs

	intervention effect (RR)													
baseline effect	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.65	0.7	0.75	0.8	0.85	0.9	0.95
0.1	£192,889	£128,593	£96,444	£77,156	£64,296	£55,111	£48,222	£35,071	£32,148	£29,675	£27,556	£25,719	£24,111	£22,693
0.2	NA	£385,778	£192,889	£128,593	£96,444	£77,156	£64,296	£42,864	£38,578	£35,071	£32,148	£29,675	£27,556	£25,719
0.3	NA	NA	NA	£385,778	£192,889	£128,593	£96,444	£55,111	£48,222	£42,864	£38,578	£35,071	£32,148	£29,675
0.4	NA	NA	NA	NA	NA	£385,778	£192,889	£77,156	£64,296	£55,111	£48,222	£42,864	£38,578	£35,071
0.5	NA	NA	NA	NA	NA	NA	NA	£128,593	£96,444	£77,156	£64,296	£55,111	£48,222	£42,864

4. 2-way sensitivity analysis of time horizon and utility gain.

Both the time horizon and utilty gain were varied simultaneaously. The time horizon was vaired in increments of 16 weeks from 1 year to 5 years. The utility gain was varied from 0.05 to 0.12 (base case = 0.09). The results can be seen below in table. The orange cells show where the ICERs are between £20,000 and £30,000, and the green cells show where the ICERs are under £20,000. Only when the time horizon is over 3 years and with at least a utility gain similar to that of the base case is the intervention cost effective. it is however potentially unlikely

that all responders would maintain their response to the treatment into the longer term as it depends n the application of the techniques learned.

	Time horizon (weeks)													
Utility gain	52	68	84	100	116	132	148	164	180	196	212	228	244	260
0.05	£206,298	£97,879	£64,160	£47,720	£37,987	£31,552	£26,981	£23,567	£20,919	£18,807	£17,082	£15,647	£14,434	£13,396
0.06	£171,915	£89,396	£60,403	£45,610	£36,638	£30,615	£26,293	£23,040	£20,504	£18,470	£16,804	£15,413	£14,235	£13,224
0.07	£147,356	£82,266	£57,061	£43,679	£35,381	£29,733	£25,639	£22,537	£20,104	£18,145	£16,534	£15,186	£14,041	£13,057
0.08	£128,936	£76,190	£54,070	£41,904	£34,208	£28,900	£25,018	£22,055	£19,720	£17,832	£16,274	£14,966	£13,853	£12,894
0.09	£114,610	£70,949	£51,377	£40,268	£33,110	£28,112	£24,425	£21,593	£19,350	£17,529	£16,021	£14,752	£13,669	£12,734
0.1	£103,149	£66,383	£48,939	£38,755	£32,080	£27,366	£23,860	£21,150	£18,993	£17,236	£15,776	£14,544	£13,490	£12,579
0.11	£93,772	£62,369	£46,723	£37,352	£31,112	£26,659	£23,321	£20,725	£18,650	£16,952	£15,538	£14,342	£13,316	£12,428
0.12	£85,958	£58,813	£44,698	£36,047	£30,201	£25,987	£22,805	£20,317	£18,319	£16,678	£15,308	£14,145	£13,146	£12,280

5. Using alternative sources of utility data.

Results can be seen below. When the utilities from Lloyd 2011[13] are being used the ICER has fallen compared to the base case because the incremental QALY has increased, which leads to a smaller ICER. This was expected because the incremental gain in utility if a patient responds is more than in the base case.

Table 24: SA4 results (per person) – using Lloyd 2011 utilities

	Total cost	Total QALY
MPH	£0	0.7231
MPH + BT	£868	0.7332
	£868	0.0101
ICER	**£85,958**	

When using the utilities from Hodgkins 2013 [9], the results did not change compared to the base case because the incremental QALY was the same (note that sensitivity analyses were only run deterministically and so the results here match the deterministic base case results in the model, whereas the base case results reported in this report are the probabilistic). We can see however that in terms of total QALYs these are higher than in the previous table or in the base case, and that is because the baseline utility associated with no response is higher (0.809) from Hodgkins 2013.

Table 25: SA4b results (per person) – using Hodgkins 2013 utilities

	Total cost	Total QALY
MPH	£0	0.8263
MPH + BT	£868	0.8339
	£868	0.0076
ICER	**£114,610**	

1.8 MPH + self-help behavioural therapy model: Discussion

1.8.1 Summary of results

The results of this model show that in a population of children who are on MPH but have some remaining functional impairment, the addition of self-help behavioural therapy is not cost effective.

Threshold analyses have shown that the cost of the intervention would have to be significantly lower to make the addition of the behavioural therapy cost effective. Varying the effectiveness in a two way sensitivity analysis did not change the results to cost effective for any pair of baseline and treatment effect tested. Assuming the effectiveness peaks earlier (at 6 months rather than 12) improved the ICER, but still not a level that would imply the intervention is cost effective. Using utility values from different sources also had an impact on the ICER, as a larger incremental gain in utility between a responder and non-responder will lead to a larger incremental QALY and a smaller ICER. Although as the base case ICER was very high, it still remained above the NICE threshold even with a larger QALY gain.

1.8.2 Limitations and interpretation

The results have to be interpreted with caution, because the model is only comparing the addition of a self-help non-pharmacological intervention on top of what was used as a baseline in the study (on MPH). It does not tell us about what else might be cost effective that a patient could add or switch to if they are a partial responder, only that what we have investigated as an addition is not cost effective. It also needs to be interpreted with caution as to whether the results can be extrapolated to other treatments that patients might only be partially responding to. But given the 2-way sensitivity analysis, we can be fairly confident that even another treatment with a higher baseline response rate or higher treatment effect wold still not improve the ICER to a level considered cost effective.

This model is not without its limitations. It is only based on a single study. The committee thought that it would be useful to use this study as a basis for the model because the intervention was different to standard forms of behavioural therapy like parent training, and particularly because it is more longer term it was felt to be an intervention that could be a baseline form of psycho-social treatment because it is ongoing rather than a short course of treatment with no follow ups.

It can be difficult to also marry-up the conclusions of the model with what might be interpreted from the clinical review about the interventions in question. This is particularly because of the model using dichotomous outcomes whereas the clinical review is using continuous outcomes for decision making. On a continuous scale, the improvements may be more subtle and there could still be an improvement in quality of life even if someone hasn't gone from non-response to response. The guideline clinical review also had to decide on a Minimally Important Difference (MID) threshold to decide if an intervention has clinical benefit. The threshold decided upon with the GC was a difference of >20% of the control group risk. Therefore the clinical review has also used a cut-off but in a different way, because the cut-off is using the difference between the control group and the intervention group change or final scores to imply whether the intervention is better than the comparator, whereas the studies that specifically report dichotomous outcomes are looking at the difference in each group and deciding what proportion respond to each. It is possible that the dichotomous outcomes from a paper may be reaching different conclusions to what the clinical review might be reaching from the continuous outcomes. For the study this model is based on (Dose 2016), the clinical review did not find the intervention clinically effective using the cut-off of >20% of the control group risk. The clinical review also did not extract dichotomous outcomes for studies that also reported clinical outcomes, but the MID for dichotomous outcomes in the clinical review was decided as a difference of 50 people more per 1000. Finding the absolute values of numbers of responders gives an absolute value of 162 more per 1000, i.e. there will be 162 more responders per 1000 in the intervention group compared to the control group. Using the clinical review MID for dichotomous outcomes implies that the intervention has clinical benefit. Therefore the two outcomes are in conflict here.

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 particularly for behavioural therapies - the effectiveness of these on the condition are 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. The impact this may have in the model is to potentially improve

the cost effectiveness of the intervention (lower the ICER). It is however unclear if that would be to the level of making the intervention cost effective.

As the structure was kept simple without any assumptions about changes to treatment or costs of the baseline treatment, the model could have underestimated the costs. As changes to medication were allowed in the study, it is possible that the effect of drug changes has already been captured through the response rates. The costs of this being omitted could impact the model, but it is difficult to interpret what impact this may have because we do not know how the behavioural therapy may have influenced drug choices and changes, as it is the incremental impact that is key. It is believed that behavioural therapy could have an impact of medication adherence, as well as the main purpose of targeting the behaviour of the child. Therefore in theory if more people changed treatments in the comparator arm then costs will increase for the comparator because treatment changes involve more staff costs, and other treatments are also likely to be more costly than methylphenidate, overall lowering the incremental costs and the ICER. However, the study findings seem to be at odds with this however because it states “chi-square tests indicated that a significantly lower percentage of EG (enhancement group - the intervention group) children still received methylphenidate at post assessment compared to CG (control group) children, whereas significantly more EG children received Lisdexamfetamine”. This implies that more children in the intervention group changed to more expensive treatments, which would make the intervention even more expensive and even less cost effective.

The committee were also concerned about the QALY as a measure and whether this is appropriate for capturing quality of life in ADHD. There is no empirical evidence to suggest the EQ-5D is not a valid measure of health related quality of life in this area. To reassure the committee, sensitivity analyses were conducted using alternative methods and measures of capturing quality of life associated with ADHD and its changes. Differences in quality of life have an impact on the results because the incremental QALYs are very small and therefore small changes can have a large impact on the ICERs.

No deterioration has been assumed, although this may be captured by the response probabilities.

1.8.3 Generalisability to other populations or settings

It’s possible that the results could be generalisable to other drugs instead of methylphenidate, but this would really depend on how similar the response probabilities were, however two way sensitivity analysis showed that varying the baseline response probability still led to a high ICER. The behavioural therapy treatment itself may not be appropriate for some other mental health conditions because of the self-help rather than face to face focus.

Other settings in other countries can also have different attitudes towards the condition itself and towards some of the treatments like medication, particularly for children. The intensity and length of courses of behavioural therapy can be very different in different countries.

1.8.4 Comparisons with published studies

No published economic evaluations were identified since the last guideline that looked at the addition of behavioural therapy on to drug treatment. Some economic evaluations included in the last guideline were excluded in this update because of applicability and methodological limitations.

1.8.5 Conclusions

This model was attempting to answer the question of; “In children currently on MPH, is the addition of self-help behavioural therapy cost effective?”

The model showed that the addition of the behavioural therapy was not cost effective with a very high ICER. The results are very sensitive to small changes in costs or particularly QALYs. The model has limitations such as only being based on a single study, and no assumptions about further treatment being made. Therefore there remains some uncertainty around whether the intervention is cost effective.

1.8.6 Implications for future research

There remains a lack of economic evaluations looking specifically at combination treatments of pharmacological and non-pharmacological interventions. Most economic evidence tends to focus on the pharmacological interventions because this is driven by pharmaceutical companies. Ideally the models produced in this guideline will help to fuel further research looking into the added benefit of combination treatments.

1.9 Medication + CBT model: Methods

1.9.1 Model overview

1.9.1.1 Comparators

This model is focusing on an adolescent population (aged 14-18) comparing staying on medication if you are a partial responder versus adding CBT. The model is therefore interested in the added value of CBT on top of medication. The intervention involved 10 sessions of individual CBT, and two additional parent only sessions were offered.

This is based on a single study reporting outcomes at 4 months.

1.9.1.2 Population

The population are adolescents who are on a stable dose of medication for the last 2 months (medication is stated as being an FDA approved medication for ADHD), but have clinically significant symptoms as rated by a CGI-S rating of 3 or above.

1.9.1.3 Time horizon, perspective, discount rates used

The time horizon of the model is 12 months. 12 months has been used in all the models in this guideline because of a lack of long term data and the need to make further assumptions for a longer timeframe.

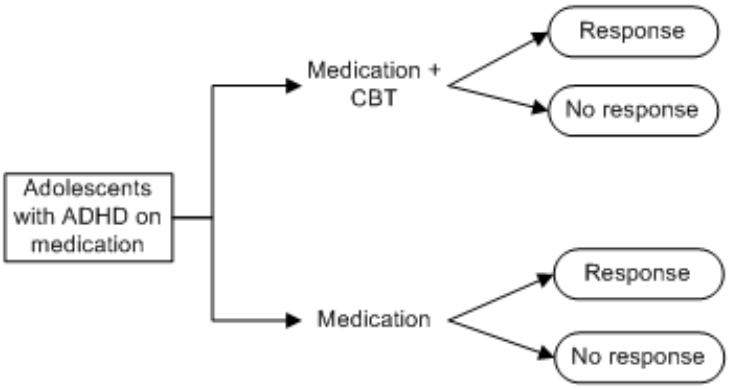
No discounting is necessary because of the time horizon of the model.

1.9.2 Approach to modelling

As with the other models, the clinical outcomes used in the model are dichotomous outcomes, as this is the only way to link to quality of life. The dichotomous outcomes are in terms of response or no response.

1.9.2.1 Model structure

The model is a decision tree model which can be seen below.



Patients who enter the model are already on medication but have some clinically significant symptoms. Patients can either stay on their medication or add CBT on top of their medication. Outcomes are in terms of response or no response at the 4 month time-point because that was the length of the trial.

The effect is extrapolated from 4 months to the end of the model (12 months). As the medication the adolescents are currently on is assumed to be the baseline or current practice, then this applies for the whole time horizon of the model. Everyone in the baseline arm of the model stays on the baseline for the whole time period regardless of whether they respond or not.

However this is only a 4 month trial, so there are advantages and disadvantages to structuring the model differently after this period.

One reason no resource use such as consultations associated with response or non-response have been included in the model is because the underlying population are all on medication, then as they would all be monitored by a clinician (psychiatrist or paediatrician) at regular intervals anyway, there would be no duplication of staff resources because of response or non-response to the behavioural therapy. It is not known however what impact the intervention might have on the underlying resource use which could lead to differences in costs between the two arms. In theory because this is a population already all on drugs, then unless the behavioural therapy helped them to not need drugs anymore, existing resource use is unlikely to be affected. It may also be that the behavioural therapy leads to better compliance to medication that then makes it more effective and then perhaps responders need less frequent consultations, and would also perhaps not be considering changing treatment which would include titration and more monitoring. However all these theories would require a lot of assumptions as to what treatments people might stop, or change on to and when, or become more adherent to, in order to include these costs. Therefore it has been assumed that resource use because they are on the drugs applies to both arms and would cancel out and has not been included. There may be some incremental costs that are not being captured; as there are fewer non-responders in the combination arm, then there are perhaps cost savings from resource use of the combination treatment (because non-responders are likely to be more expensive in the long run than responders) meaning the incremental cost is higher than it would be if these costs were included. Although this is unlikely to have a huge impact on the ICER.

Whether it is acceptable to assume the effect can be extrapolated from 4 months to 12 months in both arms is also an issue. It could be argued that if a drug that elicits only a partial response is an acceptable treatment then it might be reasonable to extrapolate this. There is no perfect response in practice and even a partial response could be considered a success. Again because this is acting as the baseline then we could interpret this as the

general population effectiveness which is what we tried to capture by having the 'other treatment' in the ATX model where there was no baseline.

As with bullet points 2 and 3 in section 1.6.2.1, it was decided to extrapolate the effects from the trial and not make further assumptions about what treatments people might go on to following the end of the trial period, as this would involve too many assumptions. As also touched on earlier – it was felt that this would be a larger omission from a model that compared a drug to a non-drug comparison directly (like the ATX model), whereas here we are interested in the addition of an intervention to a common baseline.

1.9.2.2 Uncertainty

The model was built probabilistically to take account of the uncertainty around input parameter point estimates. A probability distribution was defined for each model input parameter. When the model was run, a value for each input was randomly selected simultaneously from its respective probability distribution; mean costs and mean QALYs were calculated using these values. The model was run repeatedly – 10,000 times for the base case. Sensitivity analyses were only ran deterministically.

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 26 and in the relevant input summary tables in 1.9.3. Probability distributions in the analysis were parameterised using error estimates from data sources or assumptions where not available.

Table 26: Description of the type and properties of distributions used in the probabilistic sensitivity analysis

Parameter	Type of distribution	Properties of distribution
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	For incremental utility of responders over non-responders: 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 = $SE^2/Mean$

(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 deterministic sensitivity analyses and threshold 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.

1.9.3 Model inputs

1.9.3.1 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 28 below. More details about sources, calculations and rationale for selection can be found in the sections following this summary table.

Table 27: Summary of base-case model inputs

Input	Data	Source
Population	Adolescents with ADHD on a stable dose of MPH with some remaining clinically significant symptoms (age 14-18)	Study informing treatment effect
Time horizon	12 months	
Length of treatment	4 months	Study informing treatment effect
Treatment effect	Probability of response	Guideline clinical review

Table 28: Overview of parameters and parameter distributions used in the model

Parameter description	Point estimate	Probability distribution	Distribution parameters	Source
Probability of response from medication (baseline) [(a)]	0.182		Simulations from CODA output	Control group response probability from Sprich 2016[18]
Probability of response from from intervention (adding CBT = at 4 months) [(a)]	0.428		Simulations from CODA output	From analysis of the single study (Sprich 2016[18]) using Winbugs software.
Cost (£)				
CBT session	£97 per CBT session			PSSRU 2016 [4] From a study on depression, delivered in a CAMHS secondary care setting
Utilities				
Responder utility	0.83	Beta	Alpha = 489.7 Beta = 100.3	Van Der Kolk 2014 [20]
Non-responder utility	0.74	Beta	Alpha = 436.6 Beta = 153.4	Van Der Kolk 2014 [20]

Parameter description	Point estimate	Probability distribution	Distribution parameters	Source
Utility gain from responder over non-responder	0.09	Gamma	Alpha = 10.94 Beta = 0.008	Difference between responder and non-responder utility

Abbreviations: CAMHS; Child and Mental Health Services

(a) Note that the base case baseline and treatment effect were derived by nputting the raw values from the study into Winbugs software. The probabilistic effects were derived using CODA output from winbugs of 60,000 simulations.

1.9.3.2 Initial cohort settings

The addition of CBT to medication is compared to medication alone. The intervention is a 4 month intervention provided individually.

The quality of life of responders is assumed to increase linearly over the 4 months 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.

Those who respond are assumed to remain responders.

1.9.3.3 Baseline event rates

The baseline event rate in the model is the response probability in the medication alone arm. This is seen as the baseline because adolescents who enter the model are already on this treatment and the model is therefore investigating the additional effectiveness of adding behavioural therapy and the trade-off with the additional costs.

1.9.3.4 Relative treatment effects

The treatment effect is based on a single study (Sprich 2016) as this was the only study identified comparing medication alone with the addition of CBT in the combination question clinical review, that had the appropriate outcomes.

The study was a crossover trial, whereby initially 46 adolescents were randomised to either 4 months of CBT or 4 months of waitlist. 24 people were assigned to CBT and 22 to the waitlist control. After 4 months those who were on the waitlist were allowed to cross over and have CBT – with 19 of the waitlist group doing so. There was in effect two 4 month evaluation periods. Those who undertook CBT were in total 43 people (24 + 19), and those in the waitlist group are the 22 original waitlist patients. In total this makes 65 observation points with 4 month outcomes – not 65 patients as some of these are the same patients that had both the waitlist and then the intervention.

The outcome used to define response at 4 months was a 30% reduction in the ADHD rating scale. This is a relatively popular scale and similar cut-off scores have been used in other studies.

Both the intervention and comparator probability of response was derived by inputting the raw values from the study into Winbugs14 software. This method was undertaken so that the probability of response for the interventions being compared could be derived probabilistically using CODA output from Winbugs, which would keep the correlation between the baseline and treatment effect when draws were being taken from a simulation for the PSA. The code for the baseline and pairwise models from winbugs can be found in section 1.13.2. Fixed effects models were used because it was only one study.

1.9.3.5 Utilities

Utilities used in the model for a responder and a non-responder are from Van Der Kolk 2014 (ref). More detail on the child quality of life studies identified from a quality of life systematic search can be found in section 1.2.3.5 of the parent training model write-up. The study used the UK EQ-5D tariff, and was a fairly large sample.

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 is however no quality of life data associated with behavioural therapies for an ADHD population.

As with the previous models, sensitivity analyses were conducted using two alternative sources of utilities to see the impact this would have (see section 1.3.3.6 for an explanation of these studies).

1.9.3.6 Resource use and costs

1.9.3.6.1 Resource use of providing intervention

The table below shows the cost of the intervention.

Table 29: Intervention cost (individual CBT)

Component	Cost	Description
CBT	£97 per session * 10 sessions = £1,164	10 sessions of CBT based on PSSRU cost of a CBT session
2 optional parent only sessions	£97 per session * 2 sessions = £194	2 additional sessions for the parents only
Cost per individual without optional sessions	**= £970**	
Cost per individual with optional sessions (SA3)	**= £1,164**	

The cost of a session of CBT is from the PSSRU 2016 [4], and is based on costs estimated for a randomised controlled trial of interventions for adolescents with depression. The setting was two Child and Mental Health Services (CAMHS) teams in secondary care where CBT was delivered. This includes salary (based on the average for a specialty doctor (midpoint), clinical psychologist (band 8 median) and mental health nurse (band 6 median)), oncosts, overheads capital overheads, and the ratio of direct to indirect face to face contact..

The intervention that targeted the individual with ADHD has been included here. The inclusion of the optional parent only sessions has been included in a sensitivity analysis.

CBT in the NHS is usually individual rather than in a group because the ADHD symptoms being targeted are quite specific, and individual treatment is believed to have more benefit for targeting those core ADHD symptoms. Therefore although the intervention cost would be lower if spread over more people and provided in a group, this isn't a format that CBT is commonly provided in in the NHS, and therefore this will not be explored in a sensitivity analysis.

As mentioned previously, no cost of the underlying medication or additional resource use has been included in the model because it is assumed that this is likely to cancel out if people on both groups are on the baseline treatment, and it is uncertain what impact CBT may have on underlying resource use and drug treatment.

1.9.4 Computations

The model was a decision tree model 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.

Patients start at time zero and have the interventions for 4 months. At 4 months patients are assigned as being responders or non-responders based on effectiveness from the trial. The proportion of people that are responders to any of the treatments are applied the responder utility linearly over the 4 months to represent a slowly increasing level of benefit (through utility) from that of baseline (non-response utility) to that of a responder. Responders remain responding for the remaining time in the model. No assumptions are made about what happens to non-responders or changes to the baseline treatment either as this is assumed to be common to both arms.

Response probabilities were derived from inputting the raw numbers of responders and total people in each arm into winbugs to derive probabilities of response (as well as uncertainty which was used to derive 60,000 simulations of the response probabilities from CODA output that were then used in a PSA).

No discounting was applied because the model has a one year time horizon.

Total costs and QALYs are the sum of the costs (assumed to remain static as they are based on national sources) and QALYs in each arm.

In the probabilistic analysis, only the QALYs are probabilistic because costs will not vary. The probabilistic cost per QALY was calculated by taking the average QALY per arm from all the simulations and finding the incremental and dividing the incremental costs by this incremental QALY.

1.9.5 Sensitivity analyses

1. The added effect of CBT diminishes and is linearly decreased down to baseline from 4 to 12 months.
2. Including the cost of the optional parent sessions as well.
3. 2-way sensitivity analysis of baseline effect and treatment effect.
4. 2-way sensitivity analysis of time horizon and utility gain
5. Using alternative sources of utility data.

1.9.6 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 NGC; this included systematic checking of many of the model calculations.

1.9.7 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

When there are more than 2 comparators, as in this analysis, options must be ranked in order of increasing cost then options ruled out by dominance or extended dominance before calculating ICERs excluding these options. An option is said to be dominated, and ruled out, if another intervention is less costly and more effective. An option is said to be extendedly dominated if a combination of 2 other options would prove to be less costly and more effective.

1.9.8 Interpreting Results

NICE's report 'Social value judgements: principles for the development of NICE guidance'[15] 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.

1.10 Medication + CBT model: Results

1.10.1 Base case

The Probabilistic results can be found in the table below.

Table 30: Base case results (per person)

	Total cost	Total QALY
Medication	£0	0.7561
Medication + CBT	£970	0.7748
	£970	0.0188
ICER	**£52,080**	

The combination treatment is cost-ineffective, because the incremental effect cannot justify the incremental cost. The probabilistic results also showed that the probability of the comparator treatment being cost effective at any willingness was 1 up to a threshold of £12,000 and then very close to 1 up to £30,000.

Threshold analyses:

Threshold analysis were attempted on the treatment effect, but the results were showing that even with a baseline effect of nearly zero (all other things staying the same) the ICER would still be high. Also conducting a threshold analysis on the treatment effect (and keeping the

baseline the same) gave an outcome that meant the response probability to the combination would be more than 1, which would not be possible as we are dealing with probabilities that must be bounded by 0 and 1. Because of these results, a two way sensitivity analysis was undertaken varying both the baseline effect and the treatment effect from the combination intervention to see the impact on the ICER. See results of this under sensitivity analysis 3 in the next section.

A threshold analysis on costs showed that the cost of the intervention would have to be £375 or below to make the intervention cost effective (£20,000 per QALY). This would only equate to the cost of a between 3 and 4 sessions, or given that each session is 90 minutes (based on PSSRU costs – as the study did not specify the mength of the sessions), this would equate to almost 6 hours of CBT.

A threshold analysis on QALYs showed that this would need to be 0.0485 over the 12 month time horizon to make the intervention cost effective. Compared to the base case QALY of 0.0188 this seems like a large increase proportionally, however it is still a small value.

The time horizon was also tested – still assuming that the effect from the end of treatment is maintained. This showed that the effect from responders would have to be maintained over a time horizon of around 2 years (2 times the current model time horizon) to make the intervention cost effective, keeping all other things constant. This extrapolation of effect is of course an assumption and it is questionable whether the effect would be maintained. A 2-way sensitivity analysis was also conducted varying both the time horizon and utility gain (from responders over non-responders) to see how the interaction of these two variables affects the ICER. The results of this are reported under sensitivity analysis 4 in the next section.

1.10.2 Sensitivity analyses

1. The added effect of CBT diminishes and is linearly decreased down to baseline from 4 to 12 months.

Table 31: SA1 results (per person): assuming effect decreases

	Total cost	Total QALY
Medication	£0	0.7564
Medication + CBT	£970	0.7675
	£970	0.0111
ICER	**£87,660**	

Because the model has a short time horizon, the QALY gains are very small and so the result is very sensitive to small changes in the QALYs.

2. Including the cost of the optional parent sessions as well.

Table 32: SA3 results (per person): including optional parent sessions

	Total cost	Total QALY
Medication	£0	0.7564
Medication + CBT	£1,164	0.7751
	£1,164	0.0187
ICER	**£62,159**	

Including the parent only sessions has increased the cost which has raised the ICER.

3. 2-way sensitivity analysis of baseline effect and treatment effect.

The baseline and intervention effects that were fed into the model for the two-way analysis can be seen below. In the first column are the baseline effects used, and in the first row are the intervention effects used. In the main body of the table are the ICERs that would be derived from applying the respective baseline effect and treatment effect from that row and column. The cells that contain 'NA' mean that the intervention would be the same or less effective than that baseline.

Table 33 shows the results of the sensitivity analysis. The ICERS only go below £20,000 when the baseline risk is lower than 0.3 and the treatment risk is above 0.75. implying that there would have to be at least an increase in responders of 80% over the baseline to make the intervention cost effective (all else being equal such as the cost of the intervention)

Table 33: 2-way sensitivity analysis (effect) - ICERs

	Intervention effect (RR)													
Baseline effect	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.65	0.7	0.75	0.8	0.85	0.9	0.95
0.1	£127,374	£84,916	£63,687	£50,949	£42,458	£36,392	£31,843	£23,159	£21,229	£19,596	£18,196	£16,983.16	£15,921.72	£14,985.15
0.2	NA	£254,747	£127,374	£84,916	£63,687	£50,949	£42,458	£28,305	£25,475	£23,159	£21,229	£19,595.96	£18,196.25	£16,983.16
0.3	NA	NA	NA	£254,747	£127,374	£84,916	£63,687	£36,392	£31,843	£28,305	£25,475	£23,158.86	£21,228.96	£19,595.96
0.4	NA	NA	NA	NA	NA	£254,747	£127,374	£50,949	£42,458	£36,392	£31,843	£28,305.27	£25,474.75	£23,158.86
0.5	NA	NA	NA	NA	NA	NA	NA	£84,916	£63,687	£50,949	£42,458	£36,392.50	£31,843.43	£28,305.27

4. 2-way sensitivity analysis of time horizon and utility gain

The table below shows the 2-way sensitivity analysis varying the time horizon and utility gain (responder utility minus non-responder utility) simultaneously. The time horizon has been varied from 52 weeks (1 year) to 208 weeks (4 years), in increments of 12 weeks. The utlity gain has been varied from 0.05 to 0.12 (the base case value used in the model was 0.09). The orange highlighted cells show where the ICERs are between £20,000 and £30,000, and the green highlihghted cells show where the ICERs are below £20,000. As we would expect, we can see that the higher the utility gain of a responder, the shorter the length of time that the effect of the intervention has to be maintained for (in other words – the shorter the time horizon of the model has to be). If there are in fact benefits that the model has not been able to capture – either through benefits of CBT that the trials in the review are not capturing and therefore if they were capturing them then response would be higher anyway, or through effect not always being captured through quality of life (e.g. quality of life of parents also improving due to child improving) – then an increased effectiveness and hence higher incremental QALYs would mean benefits wouldn't have to accrue so far into the future to make the intervention cost effective.

Table 34: 2-way sensitivity analysis (time horizon and utility gain) - ICERs

	Time horizon (weeks)													
Utility gain	52	64	76	88	100	112	124	136	148	160	172	184	196	208
0.05	£93,238	£73,259	£60,331	£51,281	£44,592	£39,447	£35,366	£32,051	£29,303	£26,990	£25,015	£23,310	£21,822	£20,512
0.06	£77,698	£61,049	£50,275	£42,734	£37,160	£32,872	£29,472	£26,709	£24,420	£22,492	£20,846	£19,425	£18,185	£17,094
0.07	£66,599	£52,328	£43,093	£36,629	£31,852	£28,176	£25,262	£22,893	£20,931	£19,279	£17,868	£16,650	£15,587	£14,652
0.08	£58,274	£45,787	£37,707	£32,051	£27,870	£24,654	£22,104	£20,032	£18,315	£16,869	£15,634	£14,568	£13,639	£12,820
0.09	£51,799	£40,699	£33,517	£28,489	£24,773	£21,915	£19,648	£17,806	£16,280	£14,994	£13,897	£12,950	£12,123	£11,396
0.1	£46,619	£36,629	£30,165	£25,641	£22,296	£19,723	£17,683	£16,025	£14,652	£13,495	£12,508	£11,655	£10,911	£10,256
0.11	£42,381	£33,299	£27,423	£23,310	£20,269	£17,930	£16,076	£14,568	£13,320	£12,268	£11,371	£10,595	£9,919	£9,324
0.12	£38,849	£30,524	£25,138	£21,367	£18,580	£16,436	£14,736	£13,354	£12,210	£11,246	£10,423	£9,712	£9,092	£8,547

5. Using alternative sources of utility data.

Using the utilities from Lloyd 2011 instead of those in the base case led to the results in Table 35. As the incremental gain in utility from a responder over a non-responder was higher than that in the base case then it is expected that there would be a larger incremental QALY in this analysis because we are saying there is a larger quality of life benefit for those people who respond to the treatment. therefore the ICER is lower than in the base case but still higher than in the NICE threshold.

Table 35: SA5a results (per person) – Using Lloyd 2011 utilities

	Total cost	Total QALY
Medication	£0	0.7219
Medication + CBT	£970	0.7468
	£970	0.0250
ICER	**£38,849**	

The results of using the second alternative source of utilities of Hodgkins 2013 can be seen in the table below. Because the incremental QALY gain was the same as in the base case from a responder over a non-responder (0.09), the ICER has stayed the same as the deterministic base case.

Table 36: SA5b results (per person) – using Hodgkins 2013 utilities

	Total cost	Total QALY
Medication	£0	0.8254
Medication + CBT	970	0.8441
	£970	0.018
ICER	**£51,799**	

1.11 Medication + CBT model: Discussion

1.11.1 Summary of results

The results of this study show that in a population of adolescents who are on medication but have some remaining clinically significant symptoms, the addition of a course of individualised CBT is not cost effective.

Threshold analyses have shown that the cost of the intervention would have to be significantly lower to make the addition of the behavioural therapy cost effective. Varying the effectiveness in a two way sensitivity analysis did not change the results to cost effective for any pair of baseline and treatment effect tested. Assuming the benefit from the combination treatment, in terms of quality of life, decreases linearly when the intervention ends (at 4 months) down to zero by the end of the model made the ICER increase significantly, showing that the model is very sensitive to small QALY changes.

Using alternative sources for utility data also showed the model was sensitive to QALY changes, but again not enough to make the intervention cost effective.

1.11.2 Limitations and interpretation

The model needs to be interpreted with caution because it can only be inferred that the addition of CBT is not cost effective compared to staying on something that you are only partially responding to. It is not providing any information on what other treatments might be more cost effective. There are likely to be other treatments that are more cost effective than adding CBT.

Linking on to the limitations (which are very similar to those of the previous model); it is possible that the model has captured the effect of people switching to other underlying treatments in the model through the response rates, as people could have changed medication in the trial. Therefore it is really comparing current practice with current practice plus CBT. The fact that costs haven't been included of the current practice/comparator could be a limitation because costs might be being underestimated if there is expected to be a difference in costs between the two arms. This might be likely if say underlying treatment and resource use is affected by the CBT such as people discontinuing medication or adhering more to their medication which might improve response and also prevent you from changing to other treatments, which is less costly than having consultations to titrate new treatments. Although it is uncertain what impact BT may have on resource use, and so these additional costs were not included in the model because of too many assumptions needing to be made about what impact it might have and what treatments people might change to and how often. Therefore the structure of the model that doesn't make assumptions about what might happen after someone fails treatment is a limitation.

Additionally, no deterioration has been assumed, although this may be captured by the response probabilities.

The model is only based on a single study with a small population. Dichotomous outcomes had to be used for the model because there was no way to link quality of life to continuous outcomes, as it was not possible to define levels of severity for example to be able to model more transitionally (e.g. proportion of people going from severe to moderate or mild ADHD). Therefore there is somewhat of a discord between the data that the models use and the data that the clinical review extracted, as the committee wanted to base their decisions on effectiveness on continuous outcomes. As mentioned in the limitations section of the previous model – it may be that the improvements on a continuous scale may be more subtle and there could still be an improvement in quality of life even if someone hasn't gone from non-response to response. From the clinical review using continuous outcomes; Sprich 2016 showed that the addition of CBT to mixed medication has a clinically important benefit. This agrees with the dichotomous outcome of 236 more responders per 1000 from the intervention, using the MID for continuous outcomes that the GC decided on (50 more per 1000). So in this particular case the dichotomous and continuous outcomes are in agreement. The MID's that have been decided on by the guideline committee are arbitrary however, as would be any cut-off proposed. Even though the two outcome types agree, it still remains that even though an intervention might be effective it isn't effective enough to make it cost effective.

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 particularly for behavioural therapies - the effectiveness of these on the condition are 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. The impact this may have in the model is to potentially improve the cost effectiveness of the intervention (lower the ICER). It is however unclear if that would be to the level of making the addition of CBT cost effective.

The committee were also concerned about the QALY as a measure and whether this is appropriate for capturing quality of life in ADHD. There is no empirical evidence to suggest the EQ-5D is not a valid measure of health related quality of life in this area. To reassure the committee, sensitivity analyses were conducted using alternative methods and measures of capturing quality of life associated with ADHD and its changes. Differences in quality of life have an impact on the results because the incremental QALYs are very small and therefore small changes can have a large impact on the ICERs.

1.11.3 Generalisability to other populations or settings

Whether the results of the model can be generalised to specific drugs that people may be partially responding to (as the study the model is based on recruited participants on a variety of medications) depends on the response probabilities of individual medications, but two way sensitivity analysis showed that varying the baseline probability still did not lead to an ICER below £20,000.

With regards to other populations or settings such as other mental health conditions or other countries, it is difficult to say whether the results can be generalisable because the effect of the intervention may be different as it is probably targeting different symptoms. The intensity of the treatment may be different in different settings. Also attitudes towards the condition or the treatments in general for ADHD may be different. Combination treatments may be reserved for certain populations which may be different to how it they are used in the UK, which would affect the effectiveness.

1.11.4 Comparisons with published studies

No published economic evaluations were identified since the last guideline that looked at the addition of CBT on to drug treatment.

1.11.5 Conclusions

This model was attempting to answer the question of; “In adolescents currently on medication, is the addition of Cognitive Behavioural Therapy cost effective?”

The model showed that the addition of CBT was not cost effective with a high ICER.

The results are very sensitive to small changes in costs or particularly QALYs. The model has limitations such as it is only short term with no longer term assumptions about further treatment made, and only being based on a single study for effect. Therefore there remains some uncertainty around whether the intervention is cost effective.

1.11.6 Implications for future research

There is a lack of economic evaluations looking at combination treatments. These need to be based on trial data rather than additive probabilities of two interventions which is an incorrect assumption. More needs to be known and understood about what the impact of different treatments together on the condition is. Trials needed to inform any treatment effect in a model also need to be very clear about the populations being included in their studies i.e. whether these are drug naïve, whether they are responders, non-responders or partial

responders, as it tends to be a mix. Ideally this model will inspire further research looking at the cost effectiveness of combination treatments.

1.12 Rationale for not modelling in the adults

In the economic plan, the question on combinations of pharmacological and non-pharmacological treatments was the first modelling priority.

The committee thought it would be useful to update the previous guideline economic model in adults with ADHD, evaluating the cost effectiveness of adding (individual) CBT on top of routine care (medication). The effectiveness of adding individual CBT could be useful in helping to address ADHD-related functional impairment in people who are only partial responders to medication

The committee considered this was still a relevant question in adults, as medication is the first line intervention recommended for adults, and CBT was considered to be the most effective non-pharmacological intervention from the clinical review.

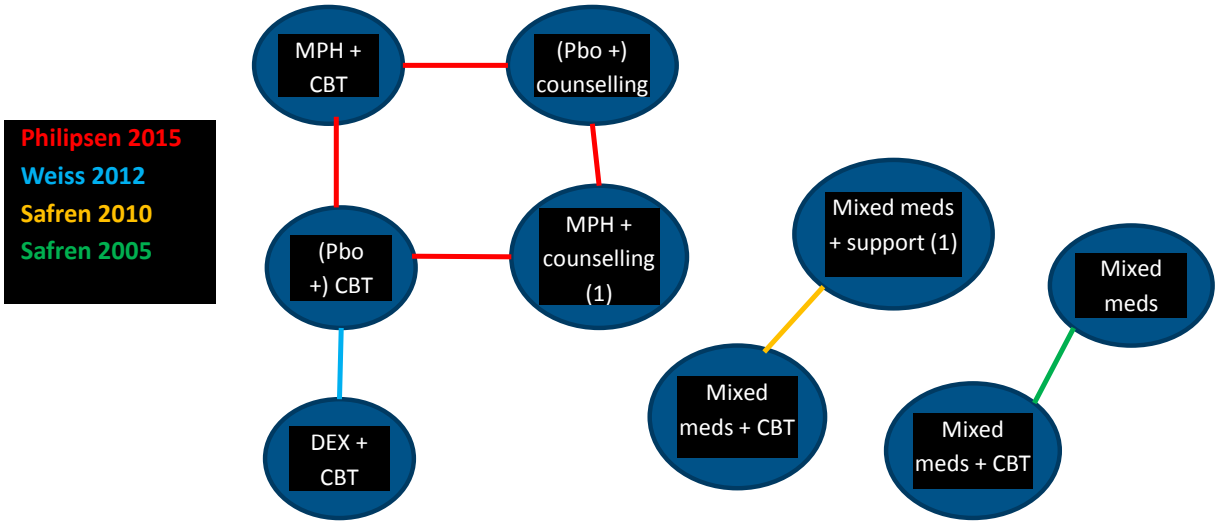
The previous model was based on a single trial in adults with a total of 31 people (Safren 2005). This study is discussed below and is included in the guideline update clinical review for this question. Therefore the previous adult model is still included in the guideline as a piece of evidence, but would be superseded by any new modelling in this area for adults.

Clinical effectiveness data

Dichotomous outcomes are the only robust method of linking to quality of life data, as there has been no further research since the previous guideline, to the health economist's knowledge, that can link changes on continuous scales to quality of life. This is very dependent on the baseline level of severity on any scale, and it has also not been possible from the clinical review perspective either to categorise studies in terms of severity.in general studies did not restrict themselves to particular populations based on severity, more commonly they were restricted either purely based on the diagnosis of ADHD or possibly based on response to previous treatment.

Out of the 10 studies that were identified from the clinical review; 4 did not have any dichotomous outcomes (Jans 2015[10], Young 2015[22], Emilsson 2011[6], Estrada 2013[7]), 2 were excluded because they were in a substance abuse population (Konstenius 2013[11], Levin 2007[12]) and it would be difficult to generalise results from this subgroup to the wider adult population. This left 4 studies that had dichotomous outcomes in the relevant population and with the relevant interventions. These 4 studies are summarised in the Table 38 at the end of the document, and Figure 4 also shows the comparisons.

Figure 4: Interventions being compared in the studies



It is difficult to combine the studies for the following reasons;

- Philipsen (2015) is the largest study and is a 4 arm trial with both a drug placebo and a control for the non-pharmacological treatment, and has multiple time points. CBT in this study was delivered as a group.
- Weiss (2012) compares dexamfetamine with (individual) CBT to a placebo with CBT.
- Safren (2010) compares adding (individual) CBT to medication in the intervention arm, versus adding relaxation with education support to medication.
- Safren (2005) compares adding (individual) CBT to medication versus medication alone.

In summary the papers are answering different questions; Philipsen (2015) and Weiss (2012) are assessing first line combination therapy, and the Safren studies assess the effect of adding another intervention to populations already stabilised on drug treatment. Although CBT is a common theme in the studies, this can be group or individual, they also use a mix of drugs, and a mix of non-pharmacological control interventions.

In addition, there is the question of how to evaluate the control groups in the studies. This is more of an issue for non-pharmacological treatments. In randomised controlled drug trials, a placebo is a non-active version of a drug to test efficacy against the active drug. Although this may not reflect reality, it is an accepted method of testing if a drug is efficacious. From the perspective of health economics, we want to see if interventions will be efficient in reality, but relying on clinical data means an assumption has to be made that if an intervention is deemed clinically effective (above its comparator, which could be placebo), then it will be considered for cost effectiveness. Placebo arms are also often considered as no treatment in an economic model. It could be argued that a placebo effect is a real effect, however a pharmacological placebo is not an option for use in the NHS. The difference with non-pharmacological 'sham' or control treatments however, is that if they are in fact effective, then this could be seen as an intervention in itself because whilst the therapist may not be following a specific intervention or programme such as CBT, talking to someone about their condition and listening to them can be a form of support that may well be used in the NHS and is advocated in many different forms already for ADHD, e.g. support groups, telephone consultations checking on patients taking their medication.

We wanted to compare medication with the addition of CBT versus medication alone, however these studies have identified some wider comparisons, particularly 'non-specific support therapy' (the non-specific counselling in the Philipsen (2015) study (called 'clinical management' in the study), and the relaxation with education support in the Safren (2010) study).

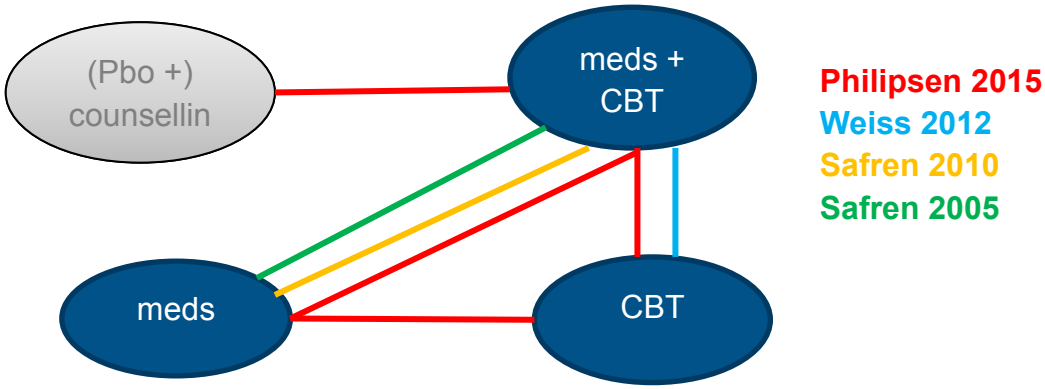
The committee discussed whether these non-specific support therapy interventions could be treated as placebos. In other words, to assume that the benefit from arms in a trial that have medication alongside a non-specific support therapy, is coming purely from the active drug intervention. Whilst this was felt not to be an issue for drug placebo, it was considered to be an assumption too far for the non-pharmacological control/non-specific treatments to be considered placebos.

However, if we were to take the view that we wanted to pool as much data as possible given the limited studies, by:

- Taking the studies descriptions at face value that they are trying to provide an attention matched control that was a non-specific therapy as a control to the non-pharmacological intervention, then we could pool all the arms that have medication plus a non-specific therapy and treat this as medication alone.
- Pooling the drugs together (which are mainly stimulants).
- Assuming a drug placebo alongside CBT is not an active treatment and treat this as CBT alone.

Then the diagram would look more like the below. We can ignore the grey comparison, as it is the two types of 'placebo' treatments and is not a comparison we are interested in for the model.

Figure 5: Pooling interventions



Also pooling treatments at similar timepoints, gives us the comparisons below at the different timepoints.

Table 37: Response probabilities from adults studies

Study	Dichotomous outcome used	Response probabilities			Time point
		Meds+ CBT	CBT	Meds	
13-15 weeks					
Philipsen 2015	decrease in CAARS ADHD index score by =>30%	0.252	0.193	0.409	13 weeks
Safren 2010	CGI scale; 2 point reduction or rated as either 1 or 2.	0.512	-	0.209	15 weeks
Safren 2005	two point change in CGI-S	0.5625	-	0.133	15 weeks
20 weeks					
Weiss 2012	much' or 'very much' improved on the CGI-I	0.652	0.154	-	20 weeks

Study	Dichotomous outcome used	Response probabilities			Time point
		Meds+ CBT	CBT	Meds	
26 weeks					
Philipsen 2015	decrease in CAARS ADHD index score by =>30%	0.355	0.229	0.455	26 weeks
52 weeks					
Philipsen 2015	decrease in CAARS ADHD index score by =>30%	0.327	0.239	0.336	52 weeks

Cells with a '-' mean that arm was not a comparator in that study.

In addition to the limitations discussed above, it is worth noting that the Philipsen (2015) study, which is the largest study (see Table 38 for numbers of people in each arm), shows that the medication arm (which was actually MPH plus non-specific therapy) has a higher proportion of responders than the medication + CBT arm, and also the CBT alone arm. This is an interesting result, suggesting that talking to someone and not providing a specific psychotherapy intervention is more effective than structured CBT. The Safren studies, that have medication in both arms, show a different picture and suggest that the combination of medication and CBT is much better than medication alone and better than medication plus relaxation. This greater effectiveness from the medication (plus non-specific therapy) arm in the Philipsen (2015) study continues at the later time points as well. It has been proposed that some of the benefit from CBT is through improving adherence to medication. This is unlikely to fully explain the difference in the results from Philipsen (2015), as both of the trial arms that have medication also have an additional non-pharmacological intervention.

If the shorter term outcomes were pooled (crudely by summing the total number of responders in each arm of each trial, and dividing this by the summed number of people in each arm of each trial), and then the proportion of responders for each time point were mapped onto a graph, we would end up with the treatment effect over time in Figure 6**Error! Reference source not found.**.

There are a number of points worth highlighting in this graph;

- At 13 weeks the response probability from the combination treatment and medication are very close together. This is because Philipsen (2015) is reducing the magnitude of difference between the combination intervention and medication intervention because its results are contrary to that of the Safren studies.
- The peak at 20 weeks from the combination treatment is because of the high effectiveness of combination versus CBT from Weiss (2010).
- At week 26, there is a change in which treatment is the most effective. This is because the only data for the later time points of 26 and 52 weeks is from Philipsen (2015).

This graph shows an erratic and inconsistent picture of the interventions effectiveness over time, this is mainly a result of the Philipsen 2015.

In summary there are not enough studies to provide a consistent picture on the effectiveness of the interventions, and excluding Philipsen 2015 would significantly reduce the number of study participants that would be informing the model.

Inputting these treatment effects into a model would mean that it is highly unlikely that a combination intervention would be cost effective. The previous model used only the effect from Safren (2005) and the MPH group had a probability of responding of 13%, and the combination group had a probability of responding of 56%. In terms of relative risk, this would be a relative risk of response of over 4 for the intervention group. Using this data, the model

found that the combination of individual CBT sessions in addition to medication was not cost effective (ICER of over £65,000). Using such a large relative risk and finding the intervention was not cost effective, leads us to conclude that inputting the treatment effect identified from the studies discussed above (and demonstrated in Figure 6) will only further confirm that combination treatments in adults are very unlikely to be cost effective. The previous model result was also assuming that response from the intervention remained at that post treatment level for the remaining time of the model. If a decrease in effectiveness over time was assumed after the course of CBT ended then the ICER would be even higher. Additionally, the cost of CBT is likely to be higher now than in the previous model, as staff costs are likely to have increased, and no staff preparation time was included, again affecting the ICER.

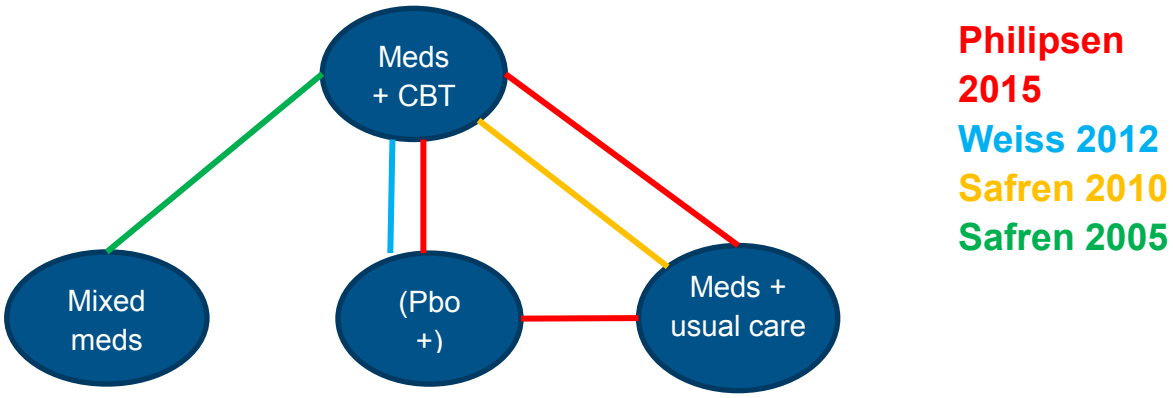
Conclusions from an updated model using this data could be challenged; if combinations were in fact cost effective, this would be driven by the smaller studies, and if combination was not cost effective, this would be driven by the larger study. Resulting in conclusions the committee would not be confident in. On this basis, the committee decided that updating the adult model would add limited value to the guideline. It would only confirm the conclusions of the previous model, that combination treatments are not cost effective in adults.

1.12.1 What is the most appropriate comparator?

What is odd from the Philipsen study, is that it implies that a drug in combination with 'non-pharmacological control' (as described in the Philipsen study) is more effective than a drug in combination with CBT.

If we accept that non-specific therapy is in fact a treatment of some kind that can have an effect, and not treat it as a placebo, then this could be a non-pharmacological type of 'usual care', as part of a holistic package of medication and some additional support. This leads to another arm in the comparison, whereby the medication plus non-specific support arms from Philipsen (2015) and Safren (2010), are considered medication plus usual care. (Figure 6).

Figure 6: Pooling interventions (assuming non-specific support is an intervention)



Economic evaluation should normally be based on pragmatic trials, since we are interested in the health effects and cost that would occur if we were to add an active non-pharmacological treatment to usual care in the NHS. Therefore there are arguments that a non-pharmacological control can miss, which are that; 'sham' or control form of psychotherapy could still have an effect, and that this 'placebo' response is a real response. One could argue that it is desirable to maximise a placebo response, as long as it is cost-effective do so. If a less structured/intervention specific approach is also using fewer resources because perhaps someone other than a psychologist could do this, or it may be provided in more of a support group based format perhaps from sectors other than the NHS

such as the voluntary sector, then it may be more cost effective than the addition of a structured course of CBT to usual care, if it is providing similar effectiveness.

It is however difficult to marry up the dichotomous outcomes extracted for any modelling, with the continuous outcomes from the clinical review. The difference in the dichotomous outcomes between interventions tends to be stronger than the difference in the continuous outcomes between interventions. It is difficult to decipher why this is, it may be because of the outcomes/scales being used and what parts of the condition they are capturing (e.g. symptoms, function, more global factors), as well as what cut-offs are being used. The threshold used to define disease status can vary between different scales and are a crude and arbitrary way of measuring improvement, as someone with very high levels of symptoms may show a significant improvement but still lie above the disease cut-off, or someone with lower levels of symptoms may show only slight improvement but then lie below the disease cut-off. Whereas continuous outcomes tend to show more consistent findings in relation to improvement following treatment. For example it is not coming across in the clinical review that stimulants + non-specific therapy is better than stimulants + CBT/DBT, in fact it found the opposite on some outcomes; i.e. that stimulants + CBT/DBT had some clinical benefit when compared to stimulants + NSST alone. Also when mixed medication + CBT/DBT was compared to mixed medication + NSST there was a clinical benefit on the CGI-I for the intervention, but not on other total ADHD outcomes. Additionally, other outcomes from the clinical review also tell us that combination treatments are better than CBT alone, and combinations tend to be better than medication alone (agreeing with the smaller studies identified with dichotomous outcomes; Safren 2005, Weiss 2012). There isn't much evidence on medication alone versus behavioural therapy alone however. Overall this is quite a confusing picture, and the dichotomous outcomes tend to confuse this even more particularly because of the Philipsen study. This also then creates a slight discord between the health economics and the clinical review because the default perspective has always been a pragmatic one of modelling the interventions that were felt to be effective based on continuous outcomes measures, as these were the primary outcome measures chosen for the review protocols.

Overall, it was felt that modelling in this area would not add any further information to what is already known; from the clinical review (which in itself is difficult to decipher because of the many outcomes and comparisons), and the conclusion reached in the previous guideline economic model in adults for this question. And if anything, could make things less clear. What we know is that combination treatments do not appear to be cost effective in adults. Previously CBT was recommended, based on consensus, in addition to medication if symptoms remained, however if there are behavioural therapy interventions that would involve fewer resources than CBT, that are considered by the committee to be similarly effective, then these might be more cost effective. This would depend however on whether these were ongoing supportive therapies, as opposed to a course for a defined period of time.

Table 38: Adults studies with dichotomous outcomes from combination review

Study	Population	Intervention	Comparator	Outcomes	Notes
Philipsen 2015	mean age = 35 Excluded if had Medication with stimulants or ADHD-specific psychotherapy within the previous 6 months before the beginning of the study.	**Intervention 1: MPH+ CBT** Received MPH (sustained release) initial dose of 10mg/day, titration with 10mg/week over 6 weeks up to 60 mg/day. Mean daily medication dosage (for both groups that took MPH I think) was 48.8mg. Twelve weekly sessions of cognitive behavioural group psychotherapy, followed by 10 monthly sessions over 52 weeks. N=107 **Intervention 2: CBT (+pbo)** Twelve weekly sessions of cognitive behavioural group psychotherapy, followed by 10 monthly sessions over 52 weeks (one every 4 weeks). Each session was in 2 parts of 50 mins each. N=109 **Intervention 3: MPH + non-specific counselling** Received MPH (sustained release) initial dose of 10mg/day, titration with 10mg/week over 6 weeks up to 60 mg/day. Mean daily medication dosage (for both groups that took MPH I think) was 48.8mg. Clinical management was the active non pharma control chosen to simulate general practice. CM participants received non-specific counselling in individual sessions (15-20 mins). Twelve weekly sessions were followed by 10 monthly sessions over 52 weeks. N=110	**non-specific counselling (+pbo)** Clinical management was the active non pharma control chosen to simulate general practice. CM participants received non-specific counselling in individual sessions (15-20 mins). Twelve weekly sessions were followed by 10 monthly sessions over 52 weeks. N=107	Decrease in CAARS ADHD index score by =>30% compared with baseline 13 weeks **Intervention 1:** 0.252 **Intervention 2:** 0.193 **Intervention 3:** 0.409 **Comparator:** 0.243 26 weeks **Intervention 1:** 0.355 **Intervention 2:** 0.229 **Intervention 3:** 0.455 **Comparator:** 0.196 52 weeks **Intervention 1:** 0.327 **Intervention 2:** 0.239 **Intervention 3:** 0.336 **Comparator:** 0.196	ITT analysis; all using those who went into the trial as denominator. Supplemental info reports total number of serious AE's and total number of AE's but only by intervention of CBT, CM, MPH, or Pbo and not in terms of the four combinations of interventions specifically.

Study	Population	Intervention	Comparator	Outcomes	Notes
Weiss 2012	Aged 18-66. Unclear about previous medication history of participants.	**Intervention: DEX +CBT** Medication was titrated in weekly increments over a 4 week phase. Dex started at 5mg and increased in increments of 5mg to a max of 20mg twice a day. Around 50% achieved this max dose. Problem Focused Therapy, *individually* for 9 sessions. Session 1 took place following titration of meds and covered psychoeducation. Patients were seen in acute treatment every two weeks (for 7 sessions) and then twice in follow up booster sessions at weeks 15 and 20. N=23	**CBT (+pbo)** Problem Focused Therapy, *individually* for 9 sessions. Session 1 took place following titration of meds and covered psychoeducation. Patients were seen in acute treatment every two weeks (for 7 sessions) and then twice in follow up booster sessions at weeks 15 and 20. N=26	much' or 'very much' improved on the CGI-I **Intervention:** 0.652 **Comparator:** 0.154 At 20 weeks	ITT
Safren 2010	Mean age 42 and 44 in the two arms respectively. Stable on medication	**Intervention: Mixed medication + CBT** No dose info or what medications specifically. 12 *individual* sessions of 50 mins each. Consisted of 3 core modules and 2 optional modules. The first module (4 sessions) focused on psycho-education including problems solving training. The second module (2 sessions) involved learning skills to reduce distractibility. The third module (3 sessions) was on cognitive restructuring. 2 optional modules were a session on application of skills to procrastination and one session including family. People for whom optional sessions were not relevant had booster sessions on prior material. The final session was on relapse and prevention. N=43	**Mixed medication + relaxation with educational support.** No dose info or what medications specifically. Most people were on stimulant monotherapy, some on stimulant duotherapy, some taking bupropion and stimulants. 12 sessions of 50 mins each, where they received info on ADHD and relaxation techniques. N=43	CGI scale; those who made a 2 point reduction or who were rated as either 1 or 2. **Intervention:** 0.512 **Comparator:** 0.209 => 30% reduction on ADHD rating scale. **Intervention:** 0.628 **Comparator:** 0.279 At 15 weeks (at posttreatment which is at 'approximately 15 weeks').	ITT (I've calculated these, as the proportion they reported were not out of all the individuals that were originally assigned to each arm)
Safren 2005	Aged 18-65 Stable on medication	**Intervention: medication + CBT** *Individual* CBT; had 3 core modules and 3 optional modules. The first core module (4 sessions) focused on psycho-education. The second core module (3 sessions) involved	**Medication** No information on what current medication they were on. N=15	Two point change in CGI-S **Intervention:** 0.562 **Comparator:** 0.133	ITT

Study	Population	Intervention	Comparator	Outcomes	Notes
		learning skills to reduce distractibility. The third core module was on cognitive restructuring. The optional modules were for people who showed evidence of clinically significant difficulties in these symptom domains (this had 3 modules of procrastination, anger and frustration, and communication). Unclear how many sessions in total. N=16		At 15 weeks	

1.13 Meta-analysis winbugs code

1.13.1 MPH + self-help behavioural therapy model

1.13.1.1 Baseline code

MPH combo model baseline (MPH alone arm)

==========

1 **trial**

==========

```
# Binomial likelihood, logit link
# Baseline fixed effects model
model{                              # *** PROGRAM STARTS
for (i in 1:ns){                    # LOOP THROUGH STUDIES
    r[i] ~ dbin(p[i],n[i])               # Likelihood
    logit(p[i]) <- m                          # Log-odds of response
#Deviance contribution
    rhat[i] <- p[i] * n[i] # expected value of the numerators
    dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))
          +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))
  }
totresdev <- sum(dev[])             # Total Residual Deviance
m ~ dnorm(0,.0001)                  # vague prior for mean
logit(R) <- m                       # posterior probability of response
}
```

Data

list(ns=1) # ns=number of studies

r[]	n[]	#	Study ID
10	52	#	1

END

Inits

list(m=0)

list(m= -1)

list(m = 1)

1.13.1.2 Pairwise meta-analysis code

MPH combo
treatment 1 = MPH
treatment 2 = Combo


```
# Binomial likelihood, logit link
# Fixed effects model
model{                              # *** PROGRAM STARTS
for(i in 1:ns){                     # LOOP THROUGH STUDIES
    mu[i] ~ dnorm(0,.0001)          # vague priors for all trial baselines
    for (k in 1:na[i])  {           # LOOP THROUGH ARMS
        r[i,k] ~ dbin(p[i,k],n[i,k])    # binomial likelihood
# model for linear predictor
        logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]]
# expected value of the numerators
        rhat[i,k] <- p[i,k] * n[i,k]
#Deviance contribution
        dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
             +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-
rhat[i,k])))
      }
# summed residual deviance contribution for this trial
    resdev[i] <- sum(dev[i,1:na[i]])
     }
totresdev <- sum(resdev[])      # Total Residual Deviance
d[1]<-0    # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
# obtain all pairwise ORs
for (c in 1:(nt-1)){
    for (k in (c+1):nt)  {
        OR[c,k] <- exp(d[k] - d[c])
        LOR[c,k]<-(d[k]-d[c])
      }
  }
# Provide estimates of treatment effects T[k] on the natural (probability)
scale
# Given a Mean Effect, meanA, for 'standard' treatment A,
# with precision (1/variance) precA
A ~ dnorm(meanA,precA)
for (k in 1:nt) { logit(T[k]) <- A + d[k]  }
}                                   # *** PROGRAM ENDS
```

Data

# ns= number of studies; nt=number of treatments
list(ns=1, nt=2, meanA=-1.474, precA=7.711764474)

r[,1]	n[,1]	r[,2]	n[,2]	t[,1]	t[,2]	na[]
10	52	18	51	1	2	2

END

Initial Values
#chain 1
list(d=c(NA, 0), mu=c(0))
#chain 2
list(d=c(NA, -1), mu=c(-3))
#chain 3
list(d=c(NA, 2), mu=c(2))

1.13.2 MPH + CBT model

1.13.2.1 Baseline code

Adolescent CBT baseline Data (drug alone arm)

==========

1 **trial**

==========

```
# Binomial likelihood, logit link
# Baseline fixed effects model
model{                          # *** PROGRAM STARTS
for (i in 1:ns){                # LOOP THROUGH STUDIES
    r[i] ~ dbin(p[i],n[i])            # Likelihood
    logit(p[i]) <- m                        # Log-odds of response
#Deviance contribution
    rhat[i] <- p[i] * n[i] # expected value of the numerators
    dev[i] <- 2 * (r[i] * (log(r[i])-log(rhat[i]))
          +  (n[i]-r[i]) * (log(n[i]-r[i]) - log(n[i]-rhat[i])))
  }
totresdev <- sum(dev[])         # Total Residual Deviance
m ~ dnorm(0,.0001)              # vague prior for mean
logit(R) <- m                   # posterior probability of response
}
```

Data

list(ns=1) # ns=number of studies

r[]	n[]	#	Study ID
4	22	#	1

END

Inits

list(m=0)

list(m= -1)

list(m = 1)

1.13.2.2 Pairwise meta-analysis code

Adoelscent CBT model
treatment 1 = dugs alone
treatment 2 = combo

This code is part of
Dias, S., Welton, N.J., Sutton, A.J. & Ades, A.E. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011; last updated September 2016 (available from http://www.nicedsu.org.uk).
This work should be cited whenever the code is used whether in its standard form or adapted.

```
# Binomial likelihood, logit link
# Fixed effects model
model{                              # *** PROGRAM STARTS
for(i in 1:ns){                     # LOOP THROUGH STUDIES
    mu[i] ~ dnorm(0,.0001)          # vague priors for all trial baselines
    for (k in 1:na[i])  {           # LOOP THROUGH ARMS
        r[i,k] ~ dbin(p[i,k],n[i,k])    # binomial likelihood
# model for linear predictor
        logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]]
# expected value of the numerators
        rhat[i,k] <- p[i,k] * n[i,k]
#Deviance contribution
        dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
             +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-
rhat[i,k])))
      }
# summed residual deviance contribution for this trial
    resdev[i] <- sum(dev[i,1:na[i]])
     }
totresdev <- sum(resdev[])          # Total Residual Deviance
d[1]<-0    # treatment effect is zero for reference treatment
# vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,.0001) }
# obtain all pairwise ORs
for (c in 1:(nt-1)){
    for (k in (c+1):nt)  {
        OR[c,k] <- exp(d[k] - d[c])
        LOR[c,k]<-(d[k]-d[c])
      }
  }
# Provide estimates of treatment effects T[k] on the natural (probability)
scale
# Given a Mean Effect, meanA, for 'standard' treatment A,
# with precision (1/variance) precA
A ~ dnorm(meanA,precA)
for (k in 1:nt) { logit(T[k]) <- A + d[k]  }
}                                   # *** PROGRAM ENDS
```

Data

# ns= number of studies; nt=number of treatments
list(ns=1, nt=2, meanA=-1.606, precA=2.924053215)

r[,1]	n[,1]	r[,2]	n[,2]	t[,1]	t[,2]	na[]
4	22	18	43	1	2	2

END

Initial Values

```
#chain 1
list(d=c( NA,0), mu=c(0))
#chain 2
list(d=c( NA,1), mu=c(-3))
#chain 3
list(d=c( NA,2), mu=c(-3))
```

References

1. BMJ Group and the Royal Pharmaceutical Society of Great Britain. British National Formulary for Children. Available from: https://www.evidence.nhs.uk/formulary/bnf/current Last accessed: 04 April 2017.

2. Clemow DB, Bushe CJ. Atomoxetine in patients with ADHD: A clinical and pharmacological review of the onset, trajectory, duration of response and implications for patients. Journal of Psychopharmacology. 2015; 29(12):1221-1230

3. 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

4. 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/

5. Dose C, Hautmann C, Buerger M, Schuermann S, Woitecki K, Doepfner M. Telephone-assisted self-help for parents of children with attention-deficit/hyperactivity disorder who have residual functional impairment despite methylphenidate treatment: a randomized controlled trial. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2017; 58(6):682-690

6. Emilsson B, Gudjonsson G, Sigurdsson JF, Baldursson G, Einarsson E, Olafsdottir H et al. Cognitive behaviour therapy in medication-treated adults with ADHD and persistent symptoms: a randomized controlled trial. BMC Psychiatry. 2011; 11:116

7. Estrada RV, Bosch R, Nogueira M, Gomez-Barros N, Valero S, Palomar G et al. Psychoeducation for adults with attention deficit hyperactivity disorder vs. cognitive behavioral group therapy: a randomized controlled pilot study. Journal of Nervous and Mental Disease. 2013; 201(10):894-900

8. 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

9. 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

10. Jans T, Jacob C, Warnke A, Zwanzger U, Gros-Lesch S, Matthies S et al. Does intensive multimodal treatment for maternal ADHD improve the efficacy of parent training for children with ADHD? A randomized controlled multicenter trial. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2015; 56(12):1298-1313

11. Konstenius M, Jayaram-Lindstrom N, Guterstam J, Philips B, Beck O, Franck J. Methylphenidate for ADHD in adults with substance dependence: A 24-week randomized placebo-controlled trial. European Psychiatry. 2013; 28(Suppl 1):1

12. Levin FR, Evans SM, Brooks DJ, Garawi F. Treatment of cocaine dependent treatment seekers with adult ADHD: double-blind comparison of methylphenidate and placebo. Drug and Alcohol Dependence. 2007; 87(1):20-29

13. 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

14. Man KKC, Coghill D, Chan EW, Lau WCY, Hollis C, Liddle E et al. Association of risk of suicide attempts with methylphenidate treatment. JAMA Psychiatry. 2017; Epublication

15. 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

16. Secnik K, Matza LS, Cottrell S, Edgell E, Tilden D, Mannix S. Health state utilities for childhood attention-deficit/hyperactivity disorder based on parent preferences in the United Kingdom. Medical Decision Making. 2005; 25(1):56-70

17. So CY, Leung PW, Hung SF. Treatment effectiveness of combined medication/behavioural treatment with Chinese ADHD children in routine practice. Behaviour Research and Therapy. 2008; 46(9):983-992

18. Sprich SE, Safren SA, Finkelstein D, Remmert JE, Hammerness P. A randomized controlled trial of cognitive behavioral therapy for ADHD in medication-treated adolescents. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2016; 57(11):1218-1226

19. Svanborg P, Thernlund G, Gustafsson PA, Hagglof B, Poole L, Kadesjo B. Efficacy and safety of atomoxetine as add-on to psychoeducation in the treatment of attention deficit/hyperactivity disorder: a randomized, double-blind, placebo-controlled study in stimulant-naive Swedish children and adolescents. European Child and Adolescent Psychiatry. 2009; 18(4):240-249

20. 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

21. Waxmonsky JG, Waschbusch DA, Pelham WE, Draganac-Cardona L, Rotella B, Ryan L. Effects of atomoxetine with and without behavior therapy on the school and home functioning of children with attention-deficit/hyperactivity disorder. Journal of Clinical Psychiatry. 2010; 71(11):1535-1551

22. Young S, Khondoker M, Emilsson B, Sigurdsson JF, Philipp-Wiegmann F, Baldursson G et al. Cognitive–behavioural therapy in medication-treated adults with attention-deficit/hyperactivity disorder and co-morbid psychopathology: A randomized controlled trial using multi-level analysis. Psychological Medicine. 2015; 45(13):2793-2804

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 39: 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