ORIGINAL ARTICLE

Is NICE infallible? A qualitative study of its assessment of treatments for attention-deficit/hyperactivity disorder (ADHD)

Michael Schlander [a,b,c]

[a] Institute for Innovation & Valuation in Health Care (InnoVal^HC); [b] Department of Public Health, Social and Preventive Medicine, Mannheim Medical Faculty, University of Heidelberg (Germany); [c] University of Applied Economic Sciences Ludwigshafen (Germany

Address for correspondence: Prof. Dr. med. Michael Schlander, MBA, InnoVal^HC, PO Box 1107, D-65741 Eschborn, Germany. Tel.: +49-6023-929589; Fax: +49-6023-929591; michael.schlander@innoval-hc.com

ABSTRACT

Background: Conclusions of the recent NICE technology appraisal of treatments for attention-deficit/hyperactivity disorder (ADHD) differ from recommendations by other Health Technology Assessment (HTA) agencies, such as the Scottish Medicines Consortium (CMS) and the Australian Pharmaceutical Benefits Advisory Committee (PBAC). NICE did not identify differences on grounds of clinical effectiveness between treatment options studied and issued technology guidance based on clinical profiles of compounds and on drug acquisition costs. The aim of the present study was to explore the robustness of NICE assessment methods when addressing a complex clinical problem such as the evaluation of ADHD treatment strategies. This robustness will be of interest to international policy-makers, given the widespread perception of NICE as a role model for the implementation of HTAs including economic evaluation.

Methods: A qualitative case study was performed to critically appraise the technology assessment report (AR) underlying NICE conclusions, including a systematic search for and analysis of relevant literature.

Results: The AR produced on behalf of NICE was found to exhibit a range of anomalies.

Search criteria were not applied consistently, and the available clinical evidence was not used optimally; selection of clinical endpoints and clinical trials for analysis were idiosyncratic. The primary cost–effectiveness model relied on six short-term studies only, and secondary extensions combined heterogeneous study designs and different clinical endpoints. Neither the distinction between efficacy and effectiveness nor the role of treatment compliance in ADHD was addressed adequately. Long-term extensions of the model were impaired by use of inappropriate discount rates and absence of consideration of long-term sequelae associated with ADHD.

Conclusion: A review of the literature strongly suggests that the NICE assessment of ADHD treatment strategies was incomplete and likely prone to bias. It is concluded that NICE did not adequately accommodate a complex clinical decision problem. Although the present qualitative case study of one assessment cannot, and was not designed to, invalidate the NICE approach to economic evaluation of healthcare programs, this observation may have potentially far-reaching implications for the generalizability of NICE-like approaches.

Introduction

The National Institute for Health and Clinical Excellence (NICE) is widely regarded as a role model for the implementation of Health Technology Assessments (HTAs), including economic evaluation. The editors of the *British Medical Journal* even suggested, 'NICE may prove to be one of Britain's greatest cultural exports'[1]. The NICE approach to economic evaluation is based on the logic of cost-effectiveness and relies on quality-adjusted life-years (QALYs) as a universal and comprehensive measure of health outcomes, which combines length and quality of life in a single index. It remains to be established, however, how well the highly standardized NICE approach can accommodate complex clinical decision problems. One such example is the choice of optimal treatment for children and adolescents with attention-deficit/hyperactivity disorder (ADHD)[2].

Scope

Attention-deficit/hyperactivity disorder (ADHD)

Accordingly the recent NICE appraisal of ADHD treatments[3] may serve as a case study to explore the performance of NICE technology appraisals in practice. ADHD is believed to represent the most common psychiatric disorder in children and adolescents[4–6] and is associated with a substantial economic burden[2], affecting individuals with ADHD and their caregivers, parents and other family members[7–9]. The economic impact of ADHD is further exacerbated by its frequent persistence into adulthood[5,6,10], thus constituting a chronic condition, and by serious long-term sequelae including poor driving abilities[11], higher risks of accidents and injuries[12–14], increased rates of tobacco, alcohol and other substance use disorders[15], more frequent antisocial behaviors[16,17] and encounters with the criminal justice system[18–21] across the lifespan, as well as relatively poor educational outcomes and lower-ranking occupational positions than controls[22].

The major forms of clinically proven treatment for ADHD are psychosocial interventions and medication management[6,23,24]. The scope of the present NICE Technology Appraisal was limited to drug treatment in children and adolescents[2,25]. The NICE Appraisal Committee found it was '*not possible to distinguish between the different [treatment] strategies on the grounds of cost-effectiveness*'[3,26], and 'accepted the importance of having a range of drug treatment options'[3,26]. Choice of medication should be influenced by clinical profiles of drugs, individual preferences, and the UK National Health Service (NH) acquisition costs, as well as a number of other factors including the presence of comorbidity, compliance issues, and potential for diversion of medication[3,26]. Thus guidance issued by NICE deviated from the assessment, which had concluded that 'the results of the economic model clearly identified an optimal treatment strategy of 1st-line dexamphetamine, 2nd-line methylphenidate immediate-release for treatment failures, followed by 3rd-line atomoxetine for repeat treatment failures', without reference to methylphenidate modified-release preparations[2,27].

NICE guidance also differs from recommendations issued by other HTA agencies and professional organizations. For instance, the Scottish Medicines Consortium (CMS) initially did not recommend atomoxetine in February 2005, apparently on grounds of the same evidence base as NICE, reasoning that the economic case for atomoxetine had not been demonstrated[28]. Following a full re-submission, it accepted atomoxetine for restricted use within the NHS Scotland only in June 2005, limiting its use to patients who do not respond to stimulants or in whom stimulants should not be given or are not tolerated[29]. The Australian Pharmaceutical Benefits Advisory Committee rejected atomoxetine 'because of unacceptable and uncertain cost-effectiveness'[30], even as a second line option after treatment failure with, or contraindications to, stimulants[31]. A group of European clinical experts reviewed the use of long-acting medications for ADHD and proposed a treatment guideline placing both dexamphetamine and atomoxetine as second-line options for patients who did not respond to, or suffered adverse effects from methylphenidate[32], although this group had been aware of the technology assessment done on behalf of NICE[27]. These discrepancies warrant further inquiry.

Objectives

The objective of the present report is to analyze the real-life performance and robustness of the methods for health technology assessments as adopted by NICE, when applied to a particularly challenging field for economic analysis[2], the assessment of treatment strategies for ADHD. The present report will provide a critique of the application of economic evaluation techniques on behalf of NICE, i.e. it will 'appraise the appraisers'[33]. It will further establish the context for a discussion of implications for healthcare policy-makers[34]. This analysis will be of international interest because of the policy relevance of NICE guidance in England and Wales, its international spill-over effects[32], as well as the asserted 'triumph of NICE', which has been said to be 'conquering the world'[1]. One should expect that the highly standardized approach of

NICE[2,35,36] would result in technology assessments consistently meeting highest quality standards, providing relevant information for stakeholders and decision-makers (which include, in the case of NICE, the appraisal committee and the clinical guideline development group concerned with the technology assessed), and being free of technical flaws, thus rising beyond the limitations frequently encountered with health economic evaluations[37–39]. Successful accommodation of the complexities of ADHD would be reassuring, while the presence of problems might draw attention to possible underlying reasons and hence areas of potential process improvement[34].

Methods

A qualitative case study was made of NICE Technology Appraisal No. 98, 'Methylphenidate, atomoxetine and dexamfetamine for attention-deficit/hyperactivity disorder (ADHD) in children and adolescents (Review of Technology Appraisal 13)', published March 2006[3]. The analysis presented here is part of a more comprehensive study of the ADHD appraisal by this author[2,40] and is primarily concerned with the Technology Assessment Report[27], since this document 'is used as the basis of the appraisal'[35]. The resulting critique will be presented in a spirit of scientific inquiry, and is not intended to put blame for potential shortcomings on any of the parties involved in the assessment. Rather the author's interpretation of underlying problem areas and suggestions for international policy-makers will be offered in a subsequent Commentary[34].

The study had descriptive, explorative, and explanatory elements. First, the initial phase of the study consisted of defining a theoretical framework for analysis. This included a description of NICE technology appraisal processes, which took place in a period of substantial upgrade and definition of the so-called 'reference case' analysis by NICE[2,35,36]. During this phase, a thematic framework was defined, comprising use of the 'accountability for reasonableness' concept as a process benchmark[2,41,42], the present critique of the technology assessment report underlying the appraisal, as well as a review of the clinical and economic literature on attention-deficit/hyperactivity disorder[40] in order to incorporate the complex interrelated issues involved in this technology appraisal[2].

The second phase of the study comprised data collection employing a number of closely related strategies. (1) From May 2004 to publication of guidance in March 2006, the NICE website (www.nice.org.uk) was visited at intervals of less than 1 month and checked for newly posted information and documents (including meeting minutes and announcements) on (a) the technology appraisal process and related methods, (b) clinical guideline development, (c) deliberations of the NICE Citizens' Council, and (d) ADHD. (2) Scientific articles cited in these documents were obtained for analysis. (3) Independent literature searches (using the PubMed and EBSCO databases as well as Google Scholar) were conducted for articles on ADHD diagnosis, treatment, compliance, cost, and cost-effectiveness, and were (4) complemented by a search for relevant abstracts presented at international meetings in the fields of clinical psychiatry, child and adolescent psychiatry, pediatrics, health economics, and pharmacoeconomics. All searches for literature fully covered the technology assessment period (from June to December 2004, cf. below). After May 2005, no further systematic searches for scientific literature were conducted, and new papers were added to the database in an opportunistic manner only. Collected documents were indexed using categories including study type, product tested, and subject matter (e.g., treatment compliance) for further analysis and interpretation.

The present report is primarily concerned with the use of clinical evidence for assessment[27], which was subjected to a critical appraisal by this author. This included an examination of design choices and justifications provided by the assessment group for internal and external consistency[27,34,40]. Unless specified otherwise, the following citations will refer to the assessment report[27] (AR).

Results

A detailed discussion of the appraisal process has been provided earlier[2]. The assessment protocol[43] had been completed June 22, 2004, and the assessment report[27] (AR) was prepared during the second half of the same year. This document was completed by the assessment group in December 2004 and comprised 605 pages including 13 appendices. It recapitulated the scope of the assessment[2,25] and delineated briefly the background of the health problem underlying the assessment, identifying issues related to prevalence, etiology, diagnostic criteria, symptoms, as well as psychiatric comorbidity and social impairment, but not long-term sequelae of the disorder (AR, pp. 34ff.). A brief description of the medications studied was followed by a methods section, which covered search, data extraction, and analysis strategies for the effectiveness and the cost-effectiveness reviews. Search criteria were designed broadly to identify 'ongoing and recently completed research' (AR, p. 42) by including, among others, 'conference proceedings, reports, dissertations and other grey literature' (assessment protocol[43], pp. 2ff.;

AR, pp. 41ff.); 'economic evaluations could include cost–consequence, cost–utility, cost–effectiveness analysis, cost–minimization and cost–benefit analyses' (AR, p. 50). Deviating from the assessment protocol, a restriction was introduced insofar as clinical studies were excluded if they had been only published as abstracts or as conference presentations (AR, p. 46). Likewise 'economic evaluations reported as conference proceedings or abstracts were excluded since the data they contain *may* not be complete' (AR, p. 50, italics added; for related consistency issues, see below and Commentary[34]). In a recent overview, which focused on the use of clinical data in health technology assessments of rapidly evolving technologies, all but two assessment groups producing evaluations on behalf of NICE were reported to use data from conference abstracts and presentations, and it was suggested that these 'technology assessment teams should increase their efforts to obtain further study details by contacting trialists'[44].

Following an effectiveness review, the assessment report offered 'a systematic review of the health-related quality of life and cost–effectiveness literature' (AR, p. 50, pp. 177ff.) and a critical review of three submissions by manufacturers of products evaluated (AR, pp. 192ff.). For economic modeling, efficacy data were synthesized using advanced mixed treatment comparison methods, and utility, resource utilization, and cost data as well as assumptions were described extensively. The primary model was enhanced by a number of probabilistic sensitivity analyses designed to integrate various effectiveness measures, alternative sources of utility data and different time horizons (from age 1 year up to age 18). Although the assessment group discussed limitations of its model, notably data deficiencies, the interpretation of the license of dexamphetamine, and situations where a midday dose of medication might be unworkable, it concluded that its 'evaluation clearly identified an optimal treatment strategy' (cf. AR, pp. 260ff.). The assessment report was subsequently published as a peer-reviewed contribution to *Health Technology Assessment*, apparently unchanged[45], although the NICE appeal panel had expressed disappointment about the omission of an important clinical study (referred to as 'LYBI'[46]) in the assessment report[2,40,47]. The appeal panel had noted that its 'disappointment was increased by the fact that the original [assessment] protocol had stipulated that both published and unpublished data be included'[47].

The present review of the technology assessment identified gaps of the assessment related to critical issues in the areas of scoping, the selection of clinical evidence for evaluation, the distinction between efficacy and effectiveness, including the role of treatment compliance, the methodology used to synthesize data from multiple sources, the structure of the economic model developed by the NICE assessment group, and the relationship between cost–utility findings of the assessment group and published cost–effectiveness evaluations.

Scoping

The scope defined by NICE[2,25] provided the framework for the analyses commissioned by the assessment group and was narrower than that subsequently used for development of clinical guidelines[2,48,49]. It is especially notable that the role of psychological interventions remained beyond the scope of the technology appraisal, despite their importance in clinical practice. The traditional view in Europe has been that behavioral treatment should be initiated preferably prior to pharmacotherapy[50], although in 2004 an upgrade of the European clinical guidelines for hyperkinetic disorder (which corresponds best to the impaired combined subtype of ADHD[2,51,52]) was published on behalf of the European Society for Child and Adolescent Psychiatry (ESCAP), recommending that medication should be considered when psychosocial treatments alone are insufficient[53]. Still, a key attribute of the treatment paradigm for ADHD in clinical practice remains the need to decide on the appropriate sequence of specific therapies following diagnosis, education, advice, and support[54]. To be optimally relevant, therefore, a health technology assessment of therapeutic interventions for ADHD might be reasonably expected to address the choice between behavioral treatment, medication management (including the specific type of drug to select), or the combination of both, by providing current data on their relative cost-effectiveness (cf. Discussion, below).

Data selection for assessment

The number and heterogeneity of outcome measures used in ADHD trials[55,56] presents a real challenge to a comprehensive research synthesis of treatment effectiveness. Accordingly, the Canadian authors of a recent review of 14 'long-term' studies, in which treatment was administered for 12 weeks or more, refrained from conducting a formal meta-analysis[57]. Similarly, the authors of a prior systematic review for the US Agency for Healthcare Research and Quality (AHRQ) of November 1999 also deemed quantitative meta-analysis, on the basis of 77 randomized clinical trials (RCTs) selected, 'inappropriate, since associated with a greater chance of obtaining imprecise and potentially misleading results'[58].

The general approach, as well as the specific parameters chosen for analysis of ADHD outcomes,

require careful consideration in terms of the reliability and validity of measurement instruments, since the consistency of outcome measures is particularly challenging in this condition. It is widely accepted among child and adolescent psychiatrists that parents and teachers (as used as sources for the Conners CPRS and CTRS rating scales) are the optimal informants about the symptoms and behavioral problems associated with ADHD[59], and that the Conners Ratings Scales (CRS) as a group represent the most widely used and empirically supported instrument to assess symptoms related to ADHD[55,60]. Accordingly, in a quantitative systematic review commissioned by the Canadian Coordinating Office for Health Technology Assessments (CCOHTA) in 1998[61,62], 24 out of 26 eligible studies used a Conners Scale. Furthermore, in the AHRQ Evidence Report of November 1999, the Conners' Scales were the most frequently used instruments in 78 studies selected for review[58]. The Conners scores were also, in the opinion of Schachar and colleagues (2002) who rejected a formal meta-analysis, the only scores allowing descriptive quantitative synthesis of long-term data[57].

Thus, albeit excluding measures of inattention and impulsivity, the choice of parent and teacher rating scales of hyperactivity for the effectiveness review was eminently justifiable[27]. Yet, inspection of the clinical studies selected for technology assessment reveals two important anomalies. One is an apparent consistency problem resulting from the interpretation by the assessment group of the inclusion requirement that 'studies must be of at least 3 weeks' duration'[27,43]. To make sense out of this criterion, one would expect minimum *treatment* duration of 3 weeks. Such an expectation would be consistent with the rationale for this 3 weeks' cut-off given by the assessment group, namely that 'the effect of medication on behaviour is often (not always) apparent immediately, but the impact on the social adjustment of the child may well not be apparent in the first days of therapy' (AR, p. 45). This was justified by the assessment group by way of reference to the DSM-IV diagnostic manual (AR, p. 44f.) However, despite this rationale, a minimum of 3 weeks *study* (not *treatment*) duration was used by the assessment group as the inclusion criterion. As a consequence, more than one-third of the 64 (65, including the important MTA Study[63,64], please see Discussion below) randomized trials selected for the clinical effectiveness review (cf. Figure 1) were crossover studies with observation periods shorter than 3 weeks per treatment arm (usually 5–7 days; indeed some studies specified daily crossovers between treatment modalities)[40]. Moreover, some of these crossover studies had been conducted without washout phases between treatment periods, which obscured

transparent statistical controls for potential carryover effects[40]. One might argue that this interpretation was formally correct, but it was hardly consistent with the assessment group's own reasoning that studies 'based either on single-dose administration or on treatment over a few days' had been 'carried out to clarify the mode of action […] rather than as therapeutic trials, so should not be included in assessments of clinical value' (AR, p. 45). No doubt the methodology actually applied was inappropriate for examining the clinical question raised, namely that of social adjustment.

While many very-short-term studies were included, at the same time high-quality double-blind trials with parallel-group design and 2 weeks' treatment duration were excluded from the effectiveness review (e.g., Biederman *et al.*, 2003[65]), although these exclusions were formally consistent with the predefined protocol[43]. Other RCTs were overlooked, too. This is a crucial oversight because some of these trials fulfilled the inclusion criteria for effectiveness (and cost-effectiveness) review; one (a head-to-head comparison of two of the therapeutic options considered) followed a double-blind, double-dummy parallel-group design with 6 weeks' treatment duration (Newcorn *et al.*, 2004[46], 2005[66]); another one was a placebo-controlled, dose–response study involving 297 randomized patients[67]. Therefore, the clinical effectiveness review, a major component of an assessment setting the stage for economic evaluation, was impaired by technical errors, notably an incomplete search and an inappropriate interpretation of the inclusion criterion for studies to document a minimum treatment period of 3 weeks. These problems led to an idiosyncratic selection of clinical evidence. As will be shown later, the problem of overlooked data was not limited to the clinical effectiveness review but extended to the review of economic evaluations.

Further to this, the cost–effectiveness evaluation provided by the assessment group deviated from the approach taken for the effectiveness review in a number of important ways (for justifications offered and related consistency issues, cf. Table A3 of Appendix published online). First, it relied for its 'base case' analysis on the clinician-rated Clinical Global Impression-Improvement subscale (CGI-I), scores of which were transformed into 'response rates'. *Secondary* economic analyses were performed from response rates using efficacy data from the clinician-rated Clinical Global Impression-Severity subscale (CGI-S), the parent-rated ADHD-RS, and finally the SNAP-IV scale, but again not the Conners Scales. Second, the studies chosen as inputs for cost–effectiveness analysis differed from those selected for the effectiveness review. Although, abstracting from some technical peculiarities mentioned earlier, the selection of studies for the effectiveness

review was made in a transparent manner using a set of specified quality criteria, this did *not* hold for the selection of studies used for economic modeling.

In order to calculate costs per QALY gained for the present economic model, response rates were preferred since they facilitate dichotomizing the effectiveness data on the grounds that they indicate an 'explicitly identified clinically meaningful change' (AR, p. 224). (Needless to say, dichotomization of continuous variables *per se* results in an upward distortion in variation and may impede the detection of differences between interventions under study[68].) Although the assessment group recognized that 'the choice of outcome measure is a critical design issue' (AR, p. 178) of an economic analysis, no reference at all was made to the extensive body of scientific literature[2,40,55,56] concerning the psychometric properties, i.e. the performance characteristics of the measures used. This might have prevented the pooling of clinical global impressions – assumed to capture health-related quality of life (HRQoL) – with response rates derived from narrow-band symptom scales, such as the ADHD-RS and SNAP-IV instruments (cf. AR, p. 225), which do not qualify as disease-specific HRQoL instruments as implied by the assessment group (AR, p. 178). Further this might have revealed the dubious psychometric properties of the CGI-I subscale, which consists of one item only and, by design, does not provide normative information independent of baseline level[69–72] (for justifications given by assessment group for its rejection of Conners' ratings, see AR, pp. 185f. and p. 224, and Appendix, Table A3). These data were combined with utility estimates based on parent proxy ratings, which – except for one sensitivity analysis – were derived from a study using EQ-5D questionnaires, which were completed by parents in consultation with their child[73] – contrary to a statement in the assessment report claiming standard gamble experiments as their source (AR, p. 235). None of the hypothetical health states used for proxy-rating were congruent with any one of the instruments used to determine 'responders'[27,40,73,74].

Though there is no compelling evidence to support the choice of the CGI-I as a primary outcome measure for cost–effectiveness evaluation in children and adolescents with ADHD (in fact, the assessment protocol[43] had mentioned that '*in addition* physician ratings of clinical global impression will be examined', suggesting CGI scores, not CGI-I subscale ratings, might be used to support findings), a direct consequence of these selection criteria was the substantial reduction of the evidence base available for economic analyses. After application of the selection criteria, notably including reports of CGI-I subscores, only five studies out of 65 used in the clinical effectiveness review were left for inclusion in the economic analysis (Figure 1).

The greatest data attrition was in studies involving dexamphetamine: though 13 studies with a total of 334 patients had been integrated in the effectiveness review (of which, seven studies with 221 patients reported Conners ratings), no effectiveness data remained after application of both filters. The assessment group addressed this problem of complete data absence for dexamphetamine by recurring to a study[75] involving 32 girls in a cross-over design that had been eliminated from the effectiveness review earlier in the selection, on grounds of 'inadequate data presentation' (AR, p. 338). The assessment report accounts for this anomaly by stating, 'A number of studies excluded from the effectiveness review, for reasons of data presentation, were nevertheless found to provide information on response rate. These studies were therefore included in the calculation of response rates for the cost–effectiveness analysis' (AR, pp. 225f.). It remains entirely unclear which studies in addition to this dexamphetamine trial[75] might have been added to the database. The ultimate inclusion of this particular study is noteworthy because, in the absence of other data on dexamphetamine, it drove both the efficacy synthesis as well as the withdrawal rate assumptions (cf. below, '*Economic model*' and Figure 2) underlying the conclusion of the economic modeling exercises undertaken that an ADHD treatment strategy starting with first line dexamphetamine was optimal. All study subjects were girls (see Appendix, Table A1), which is an important consideration as gender differences in ADHD are well documented[40,76] and influenced by referral bias in some studies[77]. Interestingly, the authors of the study themselves concluded that their data 'provide additional support for the usual clinical practice of beginning with methylphenidate'[75].

The studies ultimately selected for the primary economic model collectively comprised 1958 patients (with one open-label study[78,79] contributing 1323 of these patients); 1727 of whom had been observed for the minimum period of 3 weeks only. No clinical effectiveness data beyond 8 weeks' treatment duration were available in this group of studies (Table A1).

Efficacy, effectiveness, and treatment compliance

The distinction between *efficacy* (typically measured in RCTs) and *effectiveness* (real-world outcomes associated with an intervention) has long been recognized. Whereas RCTs follow an explanatory orientation ('*can* the intervention work?'), economic evaluations to be meaningful require a pragmatic orientation ('*does* the intervention work?'[80,81]). It is commonly accepted that the high internal validity of RCTs is achieved at the expense of their external validity (i.e., generalizability),

Evidence base Literature search[27]: 2908 publication titles identified and screened
AHRQ Review (Jadad *et al.*, Nov. 1999[58]): 78 trials (77 RCTs) selected
CCOHTA Review (Miller *et al.*, Dec. 1998[61]): 26 trials selected (n ~ 1000)
Schachar *et al.* (2002[57]): 14 trials ($\geq$ 12 weeks) selected (n = 1,379)
MTA Cooperative Study Group (1999[63,64]): 4 groups, 2 years, n = 579
Abikoff *et al.* (2004[153]); Klein *et al.* (2004[154]): 3 groups, 2 years, n = 103

Filter 1
RCTs examining MPH, DEX, or ATX,
alone or in combination, with or without NDT;
patients age < 18y; $\geq$ 3 weeks study duration;
reporting core symptoms, quality of life,
adverse effects, or educational performance

Effectiveness review
Focus on hyperactivity ratings

64 randomized clinical studies (n ~ 7000)
plus
NIMH MTA Study[63,64,121] (n = 435 out of n = 579)

Filter 2
Availability
of
CGI-I scores
(subscale)

Economic model
Focus on CGI-I scores

5 clinical studies (n = 1926), treatment duration 3–8 weeks,
hereof 1 study with n = 1323 (Kemner *et al.*, 2004[78,79])
and 1 study "CIC"; ?
plus
1 cross-over study previously excluded, n = 32 (Sharp *et al.*, 1999[75])

Secondary extensions
Availability
of **CGI-S**
or **ADHD-RS**
or **SNAP-IV**

Extended economic model
Focus on "response rates"
defined by four different scales

13 clinical studies ($n \geq$ 2768); 4 studies "CIC",
one "CIC" study could not be identified
plus
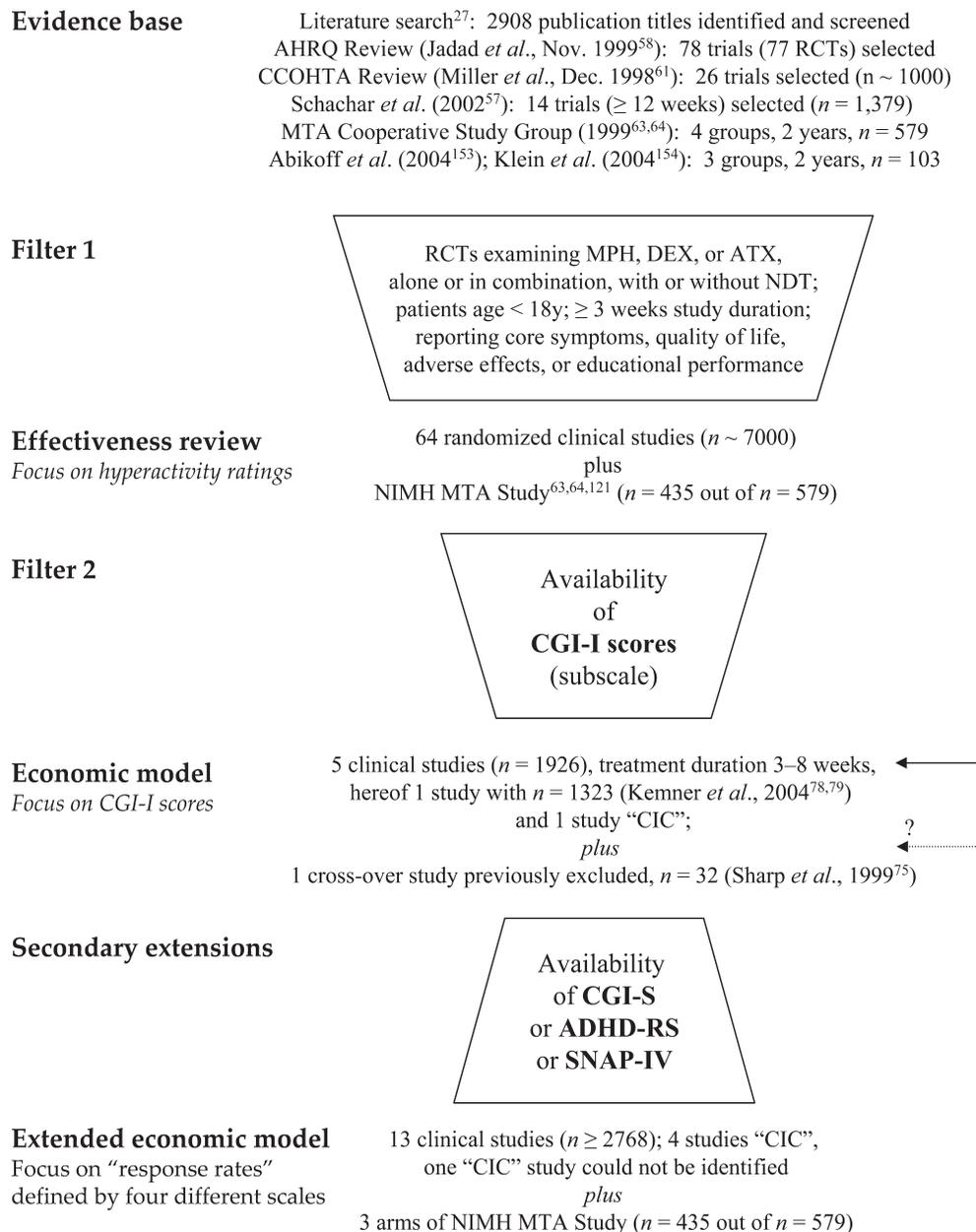3 arms of NIMH MTA Study (n = 435 out of n = 579)

Figure 1. Reduction of clinical evidence available for economic modeling after application of filters for effectiveness review and cost–utility model. CIC, commercial-in-confidence

the reason being, besides other issues such as patient and investigator selection effects, careful monitoring of study subjects designed 'to 'control' the environment … under a strict research protocol'[82], as noted elsewhere by the senior author of the assessment report: 'Great efforts are typically made in the conduct of a clinical trial to ensure that patients consume their prescribed medications. To the extent that patients do not comply with the prescribed therapy, there may be a dilution of the treatment effect originally observed in the trial'[83]. As a consequence there has been a call for more pragmatic clinical trials with minimal quality assurance and study management in psychiatry, intended to provide generalizable answers to important clinical questions without bias[84]. Accordingly, members of a recent task force initiated by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) agreed that 'it is generally acknowledged that pragmatic effectiveness trials are the best vehicle for economic studies', and expressed the view that 'artificially enhanced compliance' in RCTs is a threat to their external validity[85].

Approach by assessment group

Instead of addressing these issues with respect to their relevance for ADHD treatment, the assessment group reasoned: 'The exploration of the effects of non-

compliance would involve a number of assumptions: the assumption that RCT data capture *none* of the effects of compliance; the application of a selected estimate of compliance from a source outside of the clinical trials; and an assumption regarding the distribution of reduced compliance between morning, lunchtime and evening doses of medication. It was *felt* that these modelling assumptions would not be reasonable given the lack of available data, which would render the results of any sensitivity analysis around compliance uninformative to decision-makers' (AR, p. 233). Apparently, there was a prevailing belief that compliance would be adequately captured in controlled clinical trials and this is evident from statements in the assessment report that 'intention-to-treat analyses are favoured in assessments as they mirror the noncompliance … that [is] likely to occur when the intervention is used in practice' (AR, p. 28), and 'in our base case analysis it is assumed that the trial data adequately captures the effect of compliance on response to treatment' (AR, p. 232). This is in striking contrast not only to textbook statements mentioned earlier[82,83], which were made by the senior author of the assessment, but also to the conclusion reached in a recent review on the subject of noncompliance: 'A prime reason for the difference between efficacy in RCTs and effectiveness in the real world is the difference in patient compliance which is generally better in the context of controlled clinical trials'[86]. Specifically, in intent-to-treat analyses the clinical data of the last patient visit are carried forward for endpoint analysis, and this practice of preserving data cannot be expected to reflect the situation of a noncompliant, discontinued patient at the time when the study has been completed. Hence intent-to-treat analyses as such may in fact conceal the impact of effectiveness variables such as treatment compliance[87,88].

This approach to the problem of treatment compliance is a major issue pervading the assessment, with potentially far-reaching implications for its conclusions, because it entails comparisons between different drug regimens with different administration schedules.

The issue of noncompliance

The clinical impact of noncompliance is dependent on the condition treated as well as the medication in question[89]. Disease-specific factors that may be expected to contribute to noncompliance[90] have not been addressed in the assessment. These factors include individual and/or parental attitudes towards (psychotropic) medication that encompass potential concerns about safety and long-term treatment, as well as social stigma, particularly in association with a midday dose in children who may become the target for schoolyard bullying (aside from the potential risk of drug diversion in schoolyards[91,92], which interestingly was associated with immediate-release stimulants only[91]), disease-defining symptoms such as inattention (including their rapid recurrence 3–4 hours after the last dose of immediate-release methylphenidate), and the presence of comorbidity – externalizing disorders such as oppositional and defiant disorder and/or internalizing ones such as anxiety and depression (Table 1).

With regard to medication, a systematic review[93] of 76 clinical studies employing electronic monitoring devices (MEMS) confirmed earlier findings about the statistically significant ($p < 0.001$ among dosing schedules[93]) inverse relationship between number of daily doses required and rate of compliance. 'Dose-taking compliance' was defined in most studies included in the review as the proportion of days in which the appropriate number of doses were taken. 'Dose-timing compliance', a measure of intake of medication within the defined time frame, was defined as within 25% of

Table 1. *Specific factors affecting compliance with attention-deficit/hyperactivity disorder (ADHD) treatment (modified after Swanson 2003[90])*

Reluctance to take medication
- Social stigma associated with taking medication for a psychiatric disorder
- Embarrassment, resulting in teasing and bullying by peers
- Parental (and/or individual) attitudes to psychostimulant medication
- Concerns over long-term safety and treatment effects

Inadequate supervision

Disorder-related factors
- Oppositional and defiant behavior
- Easy distractibility
- Poor self-control
- Coexisting depression

the dosing interval (e.g., twice daily doses should be taken 12 ± 3 hours apart). Dose-timing compliance is particularly important for drugs with a duration of action of less than 24 hours.

These data provide quantitative information about the extent that compliance may be negatively influenced by more complex dosing regimens across a variety of medical conditions (Table 2). Of note, even the use of medication event monitoring systems (MEMS) may result in underreporting of noncompliance, because there is no guarantee that opening the EM device to remove a tablet means that the dose was actually taken. It is nevertheless considered to represent the current gold standard in compliance measurement. Compliance rates revealed with MEMS are more accurate and consistently lower than those estimates generated by self-reporting by patients (or caregivers), blood-level monitoring, prescription refills, or pill counts. This in turn implies that data derived from these other measurement methods will tend to overestimate compliance[94].

The clinical relevance of noncompliance in ADHD

The issue of noncompliance is arguably more relevant in ADHD than in some other chronic diseases. The authors of the Assessment Report retreated to the position that 'none of the studies in the systematic review of compliance [note added: by Claxton *et al.*, 2001[93] (Table 2)] looked specifically at ADHD' (AR, p. 233). If anything, however, then the apparently more pronounced impact of multiple daily dosing on 'dose-timing compliance', in contrast to 'dose-taking compliance', would indicate the compound magnitude of the problem relevant to ADHD, as a delay of intake would be associated with rapid recurrence of disease-defining symptoms, which include easy distractibility, poor self-regulation, and oppositional and defiant behavior[90] – all of which are likely to exacerbate compliance problems. The underlying reason is that *therapeutic coverage* depends on the relationship between pharmacokinetics (PK) and pharmacodynamic (PD) actions. On this basis, noncompliance-forgiving drugs can be differentiated from non-forgiving drugs, the latter being characterized by clinical sequelae arising from the absence of therapeutic coverage as a consequence of missed or delayed doses[89,95]. Owing to its PK/PD relationship, methylphenidate constitutes a prototypical example of a non-forgiving compound[96–99]. This fact implies that doses administered under supervision of caregivers (i.e., in particular morning doses) can be expected to be at a much lower risk of noncompliance than a midday dose to be taken by patients themselves in school, not necessarily under adequate supervision. Plasma level troughs tend to occur at the most unstructured times of the day, such as lunchtime, recess, or during bus ride home from school[100], leaving little room for doctors to tailor the timing of administration to enhance compliance, for instance by pairing medication doses with typical family activities as advocated by Weinstein (1995)[101]. The clinical relevance of these facts is broadly endorsed by expert consensus[5,6,32,40,90,102–104], including the clinical expert who contributed to the assessment[32,54].

Empirical evidence on noncompliance in ADHD

While the review by Claxton *et al.* (2001)[93] was limited to studies employing the current gold standard of compliance measurement (i.e., MEMS), there is some empirical evidence from studies in ADHD using other methods of compliance measurement. In light of their

Table 2. *Correlation of treatment compliance (adherence) with the complexity of dosing regimen, as exemplified by the number of daily doses that need to be taken[93]. Overall, compliance declined as the number of doses increased (*$p < 0.001$* among dose schedules). The following differences of dose-taking compliance in-between dosing schedules were statistically significant: o.a.d. vs. t.i.d.,* $p < 0.008$*; o.a.d. vs. q.i.d.,* $p < 0.001$*; b.i.d. vs. q.i.d.,* $p \leq 0.001$*. For dose-taking compliance, there were too few studies for statistical comparisons*

Compliance	Systematic review of MEMS studies			
	Dose-timing compliance		Dose-taking compliance	
Dosing regimen	Mean (SD)	Range	Mean (SD)	Range
1 dose / 24 h (o.a.d.)	74% (31%)	27–89%	79% (14%)	35–97%
1 dose / 12 h (b.i.d.)	58% (23%)	22–79%	69% (15%)	38–90%
1 dose / 8 h (t.i.d.)	46% (08%)	40–55%	65% (16%)	40–91%
1 dose / 6 h (q.i.d.)	40% (n.a.)	n.a.	51% (20%)	33–81%

o.a.d, once daily administration; b.i.d., administration divided in two daily doses; t.i.d., three daily doses; q.i.d., four daily doses; MEMS, medication event monitoring system; n.a., not applicable; SD, standard deviation from the mean

methodology, these studies on psychotropic medication compliance are believed to underreport the extent of the problem in ADHD[94,105]. These data have been reviewed by Hack and Chow (2001)[105], and their key findings are summarized in Table 3. A review of related literature led these authors to suspect that, 'because compliance rates are lower for children as compared to adults and psychiatric patients as compared to medical patients' […] children with psychiatric illness may be at great risk for poor medication compliance'[105].

Sometimes, practical difficulties arise from the distinction to be made between 'adherence' and 'persistence'. In general, early discontinuation of treatment (lack of 'persistence') is a common occurrence in ADHD, and 'an intriguing but unanswered question is whether the transition from punctual to erratic compliance (i.e., non-adherence) is a precursor to discontinuation'[106], although each share a number of common features. Reduced 'adherence' can be considered a significant contributor to treatment discontinuation due to perceived lack of efficacy[106].

Recently published studies provide empirical evidence supporting these expert judgments. While one Canadian study confirmed enormous variability and often-occurring low rates of persistence with methylphenidate therapy[61,107], another Canadian survey[108] revealed that 75% of parents reported that their children (ADHD patients treated with immediate-release methylphenidate [MPH-IR] divided in three daily doses ['t.i.d.']) missed doses 'from time to time', and that 55% reported missing doses in the past 2 weeks. According to these data the third daily dose was the dose most often missed. Database analyses from the US extend these findings, consistently demonstrating higher persistence rates among patients receiving modified-release methylphenidate with a 12-hour duration of action compared to those receiving mixed amphetamine salts (MAS) or immediate-release methylphenidate[109–114]: a first analysis of administrative data from the National Managed Care Benchmark Database, covering more than 17 million insured lives, had been presented in 2003[109] and published as a full paper in 2004[110]. It identified $n = 344$ children aged 6–12 years receiving MPH-IR t.i.d. and $n = 1431$ receiving a modified-release (MR) preparation of methylphenidate (MPH) with a duration of action of 12 hours[2,40] (MPH-MR12) once daily (o.a.d.). Patients receiving MPH-MR12 were significantly less likely to discontinue (47 vs. 72% among patients receiving MPH-IR over 1 year), less likely to switch (37 vs. 59%), and more likely to persist (12 vs. 1%), with nonpersistence in this study defined as the occurrence of treatment gaps greater than 14 days[110]. Retrospective evaluations of administrative data typically do not allow differential analysis of reasons for treatment discontinuation and may be distorted by effects such as patient selection bias, and it would appear conceivable that MPH-MR12 prescriptions might be associated with higher grades of impairment, which might contribute to higher rates of chronic treatment among such patients. Thus it is remarkable that the use of MPH-MR12 in this analysis was associated with significantly fewer emergency room and general practitioner visits and with a significantly lower accident and injury rate[109,110], while these patients at the same time had a higher mean number of prior diagnoses, chronic medications, and prior total medical costs[110]. Further analyses used the same database and therefore overlapping source data[111,112]. These studies extended the findings of the first analysis on the basis of 5939 individuals aged 6 years or older, who were treated either with MPH-IR (t.i.d.; $n = 1154$) or MPH-MR12 (o.a.d.; $n = 4785$). There was again a higher number of prior diagnoses among patients receiving MPH-MR12 (3.44 vs. 2.96

Table 3. *Long-term compliance (persistence) rates in children and adolescents with ADHD treated with stimulants[105]. (Note that as yet there have been no studies in child and adolescent psychiatry using electronic monitoring devices (MEMS) for compliance measurement.)*

Authors	Medication	Compliance measurement	Number of subjects	Compliance (after __ months)
Kauffman 1981[155]	MPH (and amphetamine)	Urine testing; *pill count*	$n = 12$	67% (4¼m) *87% (4¼m)*
Firestone 1982[156]	MPH	Parent report	$n = 76$	56% (10m)]
Sleator et al. 1982[157]	Stimulants	Teacher & parent report; *child report*	$n = 52$	35% (12m) *60% (12m)*
Brown et al. 1985[158]	MPH	Pill count	$n = 30$	77% (3m)
Brown et al. 1987[159]	MPH	Pill count; *parent report*	$n = 58$	75% (3m) *88% (3m)*
Johnston and Fine 1993[160]	MPH	Verbal reports	$n = 24$	80% (3m)

for patients receiving MPH-IR, $p < 0.0001$) but no significant differences between the two groups regarding the incidence of comorbid conditions associated with ADHD[112]. Use of MPH-MR12 was associated with a mean length of treatment of 199 days (compared to 108 days for MPH-IR)[112], less hospitalizations[112], and again less emergency room visits[111]. The observation of longer treatment persistence of patients receiving MPH-MR12 compared to MPH-IR was further confirmed by two independent Medicaid claims database studies in Texas[113] and California[114], respectively.

Collectively, these data illustrate the important compliance problem associated with ADHD treatment, notably with short-acting psychostimulants. As mentioned earlier, the Assessment Report did not address this issue and its undeniable implications and sequelae in relation to a meaningful economic evaluation of alternative treatment options (cf. also below, Economic model). There are two broadly accepted approaches which are available to address the problem. These are (1) the use of models to assimilate existing information from various sources combined with appropriate sensitivity analyses, and (2) the use of information from randomized pragmatic trials capturing the 'real-world' situation[115–118]. In the context of the present assessment, this approach is relevant because such pragmatic randomized real-world effectiveness studies were available, one of which[119,120] comprised a direct comparison of MPH-IR and MPH-MR12 (see Tables 4 and 6). This raises the question of how the data from this trial were integrated for analysis.

Data synthesis across endpoints and studies

The focus on CGI-I scores as the clinical effectiveness criterion for primary economic evaluation resulted in a remaining evidence base of six studies available for analysis, involving a total of 1958 patients. One open-label study contributed a disproportionate number of patients ($n = 1323$)[78], and another trial was reinstated after previously being discounted on the grounds of quality concerns[75] (Appendex Table A1). This evidence base was both qualitatively and quantitatively insufficient to assess the relative value of six alternative interventions, i.e. atomoxetine, dexamphetamine, methylphenidate in three formulations, and a hypothetical 'do nothing' alternative represented by placebo controls – notwithstanding the heterogeneous impact of treatment intensity (including but not limited to dosing), concomitant non-drug interventions and the incidence and severity of coexistent problems, including comorbidity, peer relationships, and educational performance[2,25]. In order to broaden the data basis available for analysis, the assessment group extended its primary analyses of response rates by importing data from additional trials that reported different outcome measures, specifically CGI-S, ADHD-RS, and SNAP-IV scores. This resulted in the addition of seven trials involving 822 patients (plus an unknown number of subjects included in the commercial-in-confidence study 'Quinn *et al.* 2003'[27]) over observation periods of 3–12 weeks (Appendix, Table A2). The MTA Study[63,64,121] (cf. Discussion) provided a further 579 patients, although not all data from the MTA Study were used and the assessment report is enigmatic in this regard. While the text mentions that the community comparison arm was omitted from the analysis [AR, p. 254], Table 6.17 of the assessment report notes that the behavioral treatment arm was omitted as not relevant (AR, p. 254). The medication management arm of the study was assumed by the assessment group to represent treatment with MPH-IR (AR, p. 253; cf. Discussion).

Data were synthesized across different response criteria as described earlier, apparently without assessing potential confounding effects between outcome measures and treatments – although the assessment protocol had explicated that such data would 'only be pooled when this is statistically and clinically meaningful'[43]. Furthermore, data from RCTs determining efficacy and pragmatic 'real-world' effectiveness studies were pooled. This approach could only conceal any potential effects of improved compliance, such as a greater difference between immediate-release methylphenidate and modified-release methylphenidate in real-world situations compared to experimental settings. Providing its existence, such a greater difference would be found in (appropriately designed pragmatic) studies only. This theoretical expectation is supported by a comparison of results from the pragmatic real-life study by Steele and colleagues[119,120] with those of the meta-analysis by the assessment group (Table 4). Differences in effects were invariably greater in the real-world study than in the combined meta-analysis, which comprised predominantly data from efficacy trials.

Economic model

Like the clinical effectiveness part of the assessment report, the review of the ADHD cost–effectiveness literature again revealed significant gaps. Only five presumably relevant evaluations had been identified in the published literature, including the previous NICE Technology Appraisal[122] and the Canadian assessment of CCOHTA that had used CTRS scores as effectiveness measure[123]. The assessment did not provide any reference to key cost–effectiveness publications that were directly concerned with the

Table 4. The pooling of efficacy and effectiveness trials resulted in a compression of the 'real-world' effectiveness differences between MPH-MR12 o.a.d. and MPH-IR t.i.d. Data synthesized for assessment were predominantly efficacy data from controlled clinical studies. Data in right columns integrate various endpoints. MPH, methylphenidate; IR, immediate-release formulation; MR12, modified-release formulation with a duration of action of 12 hours[2]; NNT, number needed to treat; MTC, mixed treatment comparison (technique used by assessment group for data synthesis ('meta-analysis'), ext. MTC, secondary extensions of data synthesis using mixed treatment comparison technique. All differences between treatment arms observed by Steele et al. (2006)[120] were statistically highly significant (p < 0.001). The assessment group stated only that 'the estimated response rates were subject to large uncertainty (AR, p. 236)

	CGI-I 'Real World'	CGI-I 'MTC'	CGI-S 'Real World'	SNAP-IV 'Real World'	CGI-S and -I 'synthesized' ('CGI-I baseline') 'ext. MTC'	CGI-S and -I 'synthesized' ('CGI-S baseline') 'ext. MTC'	All 'synthesized' ('ADHD-RS baseline') 'ext. MTC'	All 'synthesized' ('CGI-I baseline') 'ext. MTC'
Data source	Steele et al. 2006[120]	AR[27], p. 236	Steele et al. 2006[120]	Steele et al. 2006[120]	AR[27], p. 253	AR[27], p. 253	AR[27] p. 255	AR[27] p. 255
Response rates								
MPH-IR	62%	68%	49%	16%	76%	53%	74%	64%
MPH-MR12	83%	75%	77%	44%	85%	65%	85%	79%
Absolute difference	21%	7%	28%	28%	9%	12%	11%	15%
Relative difference	+34%	+10%	+57%	+175%	+12%	+23%	+15%	+23%
NNT (per additional responder)	4.8	14.3	3.6	3.6	11.1	8.3	9.1	6.7

interventions evaluated. These comprised at least two US cost–effectiveness analyses based on SNAP-IV-based normalization rates from the NIMH MTA Study[63,64,121] that had been in the public domain, which reported probabilistic findings from patient-level data over 14 months both for the overall study population and for subgroups defined by comorbidity[124,125], and two cost–effectiveness models comparing modified-release and immediate-release methylphenidate from the perspectives of Canadian third party payers[126] or the NHS in the United Kingdom[127], respectively. Further studies explored willingness-to-pay for new drugs[128] or had been concerned with the cost-effectiveness of atomoxetine in Canada[129]. Had the search strategy delineated in the original protocol[43] been applied appropriately and covered relevant international economic and psychiatric conferences, the assessment group would have had a chance to identify these analyses.

The basic modular structure of the economic model developed *de novo* by the assessment group is reproduced in Figure 2 (AR, p. 223). Key data inputs were required for response rates and for withdrawal rates, which 'were calculated to include all withdrawals, regardless of the reason given' (AR, p. 230), i.e. these did result in some double-counting of non-responders, as noted by the assessment group. For dexamphetamine, the only input data for both rates came from the cross-over trial in 32 girls described earlier[75]; these data provided for a withdrawal rate of zero under dexamphetamine (cf. AR, p. 231), which after data synthesis using a mixed-treatment comparison model led to an estimate of 2% compared to a range of 8–12% for the other treatment options under study (AR, p. 236).

For modeling over a time horizon of 1 year, 38 possible treatment strategies (sequences, each of them a combinations of three model 'modules') were defined

for evaluation (AR, p. 221), which were subsequently reduced to 19 strategies for analysis, without considering combination therapy. The assessment group correctly noted that this maneuver led to underestimation of decision uncertainty associated with its model. None of these strategies accommodated a switching scenario between MPH formulations. Informed by utility values ascribed to responders and nonresponders, health outcomes were expressed in terms of quality-adjusted life-years (QALYs). Calculated QALY differences between active treatment strategies (excluding the 'no treatment' option) were generally limited to the third or fourth decimal place, and exhibited inconsistent effectiveness rankings of the strategies simulated in the model (AR, p. 237). These inconsistencies disappeared only after secondary pooling of heterogeneous endpoints, however without enabling a meaningful differentiation of strategies on grounds of their effectiveness (AR, p. 242). A further analysis attempted to extend the time horizon to 12 years (from age 6 to age 18). This extension had various limitations. On a data basis that relied on short-term, predominantly well-controlled RCTs (Appendix Tables A1 and A2), it was asserted that 'the effect of compliance on response rates […] is reflected in the model' (AR, p. 250), although no allowance was made in the model for treatment nonpersistence due to compliance problems (see AR, p. 246). Despite an explicit statement to the contrary, discount rates (AR, p. 233: 'in accordance with NICE guidance') were applied that violated NICE guidance of April 2004[27,36,43], although the final scope as well as the draft and final assessment protocols had been completed only during May and June, 2004, respectively[2]. Importantly, a time horizon of 12 years, to be meaningful, would have to address long-term sequelae associated with ADHD (cf. Introduction). Except for one sentence hidden in the
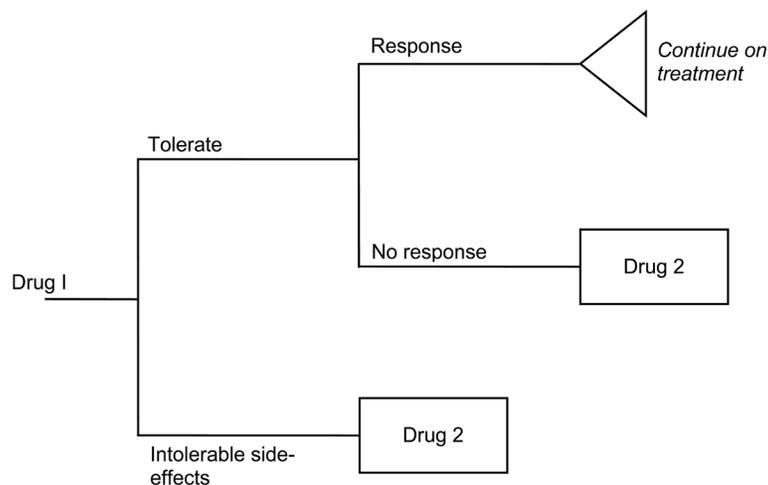

Figure 2. *Modular economic model structure. (Reproduced from King et al. 2006[45])*

results section of the assessment report (AR, p. 247), there is no hint that this issue was recognized. Current evidence of beneficial treatment effects on these sequelae is limited, and the relationship between short-term symptomatic and functional improvement and long-term outcomes has yet to be established[5,6]. Once proven, such effects might have a substantial impact on treatment cost-effectiveness. The assessment did not address this important need for further research, except for a generic caveat that 'new data on long-term outcomes could change the analysis significantly' (AR, p. 261).

Discussion

Although its scope was unnecessarily narrow, the NICE assessment even fell short of its (limited) stated objectives[25]. For instance, the evidence used did not enable analysis of the impact of diagnostic criteria and/or comorbidity on the relative cost-effectiveness of treatments under study. The economic model was driven by drug costs, since no effectiveness differences could be found. Given both the identified problems of the assessment and the differing recommendations resulting from other reviews[28–32], the question arises which further information might have been available.

A key reason for the highly selective clinical evidence was NICE's reliance on cost–utility analyses. The use of QALYs in pediatric populations however has been challenged, as there is no consensus on how quality of life should be defined and measured in children[130]. A critical review of published cost–utility analyses in child health revealed substantial variation in the methods used to calculate QALYs, with unsettling implications for comparisons across interventions for different diseases and populations[131]. Although children with ADHD were reported to experience impaired quality of life[132–134], children with ADHD tend to underestimate their disease-specific problems[59,135], especially regarding externalizing symptoms[136,137], and the validity of parent-proxy ratings is not fully understood[131]. The assessment group's assertion that 'the preferences of children and adolescents may be most relevant' and 'should be measured in patients' (AR, p. 179) thus does reflect neither the tendency of ADHD patients to underestimate their behavioral problems[59,135] nor NICE guidance specifying that 'a representative sample of the public' should be used as the 'source of preference data'[36].

The exclusive pursuit of a QALY-focused approach by the assessment group, which essentially followed NICE guidance[36], prevented the full use of important information on ADHD treatment effectiveness. As indicated earlier, a number of long-term clinical trials were available at the time of the assessment. Among these studies, the NIMH-initiated 14-months Multimodal Treatment study (MTA) is of particular relevance[63,64]. The MTA contributed 42% of the 1479 patients included in the review of long-term studies by Schachar et al. (2002)[57], and in that review it was the only trial that provided information on all 20 clinically relevant elements selected a priori for extraction. Also in the AHRQ systematic review by Jadad and colleagues (1999)[58], the MTA Study was the one trial that received the maximum quality score.

For an interpretation of the key findings of the MTA Study, it is necessary to appreciate that it was an extensively standardized, highly manualized comparison of three treatment strategies and routine community care in the United States. All four approaches tested were highly effective and showed substantial improvement from baseline at 14 months[63]. Two-thirds of the children in the community comparison group received medication, principally methylphenidate (average daily dose at study completion 22.6 mg, administered, on average, as 2.3 divided daily doses). Emphasis on subject rapport, extensive use of manuals, and regular supervision of therapists by skilled clinician investigators, together with robust monitoring measures, ensured a high degree of protocol adherence ('fidelity and compliance') for the active three treatment strategies investigated. Psychosocial interventions in the MTA Study involved three major integrated components comprising parent training, school intervention, and summer treatment program, and were designed to maximise the opportunity to demonstrate treatment effects[138,139], not cost-effectiveness. Medication management in the MTA consisted of a structured set of algorithms (starting with a double-blind, daily-switch titration protocol for methylphenidate, followed sequentially by dextroamphetamine, pemoline, and imipramine, until a satisfactory response was obtained) rather than a single medication, which like the behavioral interventions were accompanied by extensive measures to ensure protocol fidelity. Of 289 children randomized to medication management, 256 adhered to and completed the full titration protocol. Of those, 77% (198 out of 256) responded to one of the methylphenidate titration doses, and 88% (174 out of 198) were still taking methylphenidate at the end of maintenance at 14 months. Mean doses of methylphenidate at the end of 14 months were 31.1 mg per day for the combination management group and 38.1 mg per day for the medication management group ($p < 0.001$); both groups received MPH-IR divided in three daily doses ('t.i.d.')[140–142]. A wide range of outcome measures was assessed in the MTA Study, and complex relationships were observed between

parameters[143]. For instance, the presence of comorbidity was found to be an important variable influencing treatment response[144]. One of these outcome analyses, that described response rates based on averaged parent and teacher ratings of ADHD and oppositional defiant disorder symptoms on the SNAP-IV scale[121], was used for economic analyses, including the present NICE assessment. So defined response rates were 25% for the community comparison group, 34% for behavioral management, 56% for medication management, and 68% for the combination of both[121].

In light of the MTA Study design, it is noteworthy that the assessment group incorporated these results 'by assuming that the medical management group […] represents treatment with MPH-IR' (AR, p. 253). Given the administration regimens as well as the substantive efforts to manage protocol adherence in this trial, one might as well argue that the medication management arm should more appropriately have been used as a proxy for the effectiveness of modified-release methylphenidate under routine care conditions.

Economic evaluations confirmed the value of intensive medication management also in terms of its relative cost-effectiveness, with incremental cost–effectiveness ratios (ICERs) for one additional patient normalized 14 months after study entry at around US$350 for medication management (versus community care) and US$2500 for combination treatment versus behavioral treatment only[145]. For pure ADHD (i.e., ADHD according to DSM-IV diagnostic criteria, without coexisting anxiety, depression, conduct of oppositional defiant disorder), medication management dominated (i.e., it was more effective and less costly than) community care, and combination treatment versus behavioral treatment was associated with an ICER of US$940[145]. This translated into estimates of cost per QALY gained for medication management versus community care ranging between US$3000 and US$5500 (for the overall DSM-IV-defined study population; in patients with pure ADHD, medication management dominated community care), and for the comparison between combined treatment and behavioral management ranging between US$20 000 and US$40 000 for the overall study population, and US$8000 to US$15 000 for pure ADHD[125,145,146].

Subsequent extensions of these analyses[147] addressed the effects of the MTA treatment strategies on functional impairment, and revealed profound differences between patient subgroups by comorbidity: for pure ADHD, high-quality MTA-style medication management was economically superior to the studied alternatives at all levels of willingness-to-pay. For patients with coexisting conditions and at relatively higher levels of willingness-to-pay, behavioral (for the subgroup with internalizing comorbidity) and combined (for the subgroups with externalizing or both comorbidities) interventions were found likely to be cost-effective choices also[147].

The MTA Study also shed some light on the importance of treatment fidelity. Measures for treatment acceptance and attendance generally indicated very high rates of compliance over the 14-months study period, and mediator analyses showed a significant impact of treatment compliance on response rates[64]. This issue is further illuminated by two cost–effectiveness analyses[126,127] that compared immediate-release methylphenidate with a modified-release preparation with 12-hour duration of action (MPH-MR12), using Conners teacher and parent ratings as the clinical outcome measure. These analyses were extensions of the original CCOHTA model[61,123] and adopted an explicit modeling approach to analyze the impact of noncompliance (as advocated by Hughes et al., 2001[86]). Both analyses employed one- and two-way sensitivity analyses and reported, from a Canadian third-party payer perspective[126] and from the perspective of the UK NHS[127], an extended dominance of MPH-MR12 over a range of model assumptions.

The use of disease-specific outcomes criteria also provides further insights into the relative effectiveness of the treatment options under study. Relevant to the present Technology Assessment, Steinhoff and colleagues (2003)[148] presented a comparative analysis of effects sizes achieved with three once-daily ADHD medications, namely Adderall (mixed amphetamine salts, MAS, not available in Europe), atomoxetine, and a modified-release preparation of methylphenidate (MPH-MR12). These authors analyzed data from three phase III trials (with study durations of 3, 4, and 6 weeks, respectively[149–151]) used by the manufacturers of these products as part of their registration dossiers to obtain marketing authorization from the US Food and Drug Administration (FDA). All studies had enrolled a respectable number of patients, were of parallel-group, double-blind, multi-center design, and were placebo controlled. Likert-scale changes were examined on the basis of Conners ratings, and were compared using effect sizes. Effect sizes were 1.02 for MPH-MR12 and 0.62 for atomoxetine based on parent ratings, and 0.96 for MPH-MR12 and 0.44 for atomoxetine based on teacher ratings. The authors concluded that these calculations suggested that nonstimulant (atomoxetine) treatment 'is less likely to be as effective as stimulant treatment and should be positioned for trial after stimulant failure'[148]. Their results concurred with another analysis indicating an effects size of long-acting stimulants in patients with ADHD of 0.95 and that of nonstimulant medications of 0.62[152].

These findings also appear consistent with the results of two randomized head-to-head trials of MPH-MR12 versus atomoxetine[46,66,78,79], one of which had been missed in the assessment report[46,66]. Taken together, these data suggest the possibility of dominance of MPH-MR12 over atomoxetine, as the stimulant product appears at least as, or (most likely) more effective than atomoxetine, whilst being less expensive[2,40].

Summing up, aside from some peculiarities of a predominantly technical nature, the present review of the ADHD technology appraisal process revealed gaps in a number of areas of crucial clinical importance[40], concerning the role of psychosocial interventions[25], the impact of treatment noncompliance[5,80,81,86,89,90,94,95,100–103,105,106,108,118,126,127], the importance of therapeutic targets and appropriate clinical endpoint measures[55,56,59,60,69–72,135–137], the discussion of available cost–effectiveness studies[124–129], and the burden of illness resulting from caregiver and family involvement[5,7,8,12–14,18,19,109] and long-term sequelae of the disorder[10–12,17,20–22].

Collectively, the findings presented strongly suggest that the omission of critical information in the NICE technology assessment may have led to incomplete and biased results. The treatment of compliance-related issues and the absent discrimination between efficacy and effectiveness can only have worked against the long-acting medications, i.e. modified-release preparations of methylphenidate and atomoxetine. Also available data comparing methylphenidate and atomoxetine were not fully used, possibly contributing to the inability to differentiate between (long-acting) methylphenidate preparations and the non-stimulant. Potential reasons underlying these problems include a highly selective use of available evidence and apparent issues related to the internal and external consistency of the assessment (Appendix, Table A3). These observations require further evaluation, given the importance of NICE technology assessments. The discussion of underlying issues will be subject of a separate commentary[34]. To put these considerations into perspective, it is important to keep in mind that one qualitative case study cannot (and was not designed to) invalidate more than one hundred technology appraisals completed to date by NICE. It may, however, be useful as a test of the robustness of NICE technology appraisal processes. As such it may assist in identifying areas for future improvement of health technology assessment processes beyond England and Wales.

Conclusions

The NICE assessment of ADHD treatment strategies did not make optimal use of the available evidence on clinical and cost-effectiveness. Although it was restricted by its scope, it was further limited by an idiosyncratic selection of efficacy data (almost exclusively) from short-term studies, which were integrated across heterogeneous endpoints and study designs. Further technical anomalies were identified, and a review of the literature strongly suggests that the resulting assessment was incomplete and likely prone to bias. These findings indicate that NICE health technology assessments may not consistently meet expectations and therefore cast doubt on the robustness of the approach developed by NICE. While the two-stage technology appraisal process by NICE, separating assessments from appraisals, enabled NICE to moderate the putatively 'clear conclusions' of the assessment report, it could not compensate for the gaps of the technology assessment.

Acknowledgments

References

1. Smith R. The triumph of NICE. BMJ 2004;329:0
2. Schlander M. NICE accountability for reasonableness. A qualitative case study of its appraisal of treatments for attention-deficit/hyperactivity disorder (ADHD). Curr Med Res Opin 2007;23:313-21
3. National Institute for Health and Clinical Excellence [NICE]. Technology Appraisal 98: Methylphenidate, atomoxetine and dexamfetamine for attention deficit hyperactivity disorder (ADHD) in children and adolescents. Review of Technology Appraisal 13. London: NICE, March 2006
4. Faraone SV, Sergeant J, Gillberg C, Biederman J. The worldwide prevalence of ADHD: is it an American condition? World Psychiatry 2003;2:104-13
5. Wilens TE, Dodson W. A clinical perspective of attention-deficit/hyperactivity disorder into adulthood. J Clin Psychiatry 2004;65:1301-13
6. Wolraich ML, Wibbelsman CJ, Brown TE, et al. Attention-deficit/hyperactivity disorder among adolescents: a review of diagnosis, treatment, and clinical implications. Pediatrics 2005;115:1734-46
7. Leibson CL, Long KH. Economic implications of attention-deficit/hyperactivity disorder for healthcare systems. Pharmacoeconomics 2003;21:1239-62

8. Swensen AR, Birnbaum HG, Secnik K, et al. Attention-deficit/hyperactivity disorder: increased costs for patients and their families. J Am Acad Child Adolesc Psychiatry 2003;42:1415-23

9. Matza LS, Paramore C, Prasad M. A review of the economic burden of ADHD. Cost Eff Resour Alloc 2005;3:5

10. Mannuzza S, Klein RG, Moulton JL 3rd. Persistence of attention-deficit/hyperactivity disorder into adulthood: what have we learnt from the prospective follow-up studies? J Atten Disord 2003;7:93-100

11. Barkley RA. Driving impairments in teens and adults with attention-deficit/hyperactivity disorder. Psychiatr Clin North Am 2004;27:233-60

12. Swensen A, Birnbaum HG, Ben Hamadi R, et al. Incidence and costs of accidents among attention-deficit/hyperactivity disorder patients. J Adolesc Health 2004;35:346.e1-9

13. Lam LT. Attention deficit disorder and hospitalization due to injury among older adolescents in New South Wales, Australia. J Atten Disord 2002;6:77-82

14. Hoare P, Beattie T. Children with attention deficit hyperactivity disorder and attendance at hospital. Eur J Emerg Med 2003;10:98-100

15. Wilens TE, Biederman J. Alcohol, drugs, and attention-deficit/hyperactivity disorder: a model for the study of addictions in youth. J Psychopharmacol 2006;20:580-8

16. Thapar A, van den Bree M, Fowler T, et al. Predictors of antisocial behaviour in children with attention deficit hyperactivity disorder. Eur Child Adolesc Psychiatry 2006;15:118-25

17. Rasmussen P, Gillberg C. Natural outcome of ADHD with developmental coordination disorder at age 22 years: a controlled, longitudinal, community-based study. J Am Acad Child Adolesc Psychiatry 2000;39:1424-31

18. Rösler M, Retz W, Retz-Junginger P, et al. Prevalence of attention-deficit/hyperactivity disorder (ADHD) and comorbid disorders in young male prison inmates. Eur Arch Psychiatry Clin Neurosci 2004;254:365-71

19. Siponmaa L, Kristiansson M, Jonson C, et al. Juvenile and young adult mentally disordered offenders: the role of child neuropsychiatric disorders. J Am Acad Psychiatry Law 2001;29:420-6

20. Johansson P, Kerr M, Andershed H. Linking adult psychopathy with childhood hyperactivity-impulsivity-attention problems and conduct problems through retrospective self-reports. J Personal Disord 2005;19:94-101

21. Sourander A, Elonheimo H, Niemela S, et al. Childhood predictors of male criminality: a prospective population-based follow-up study from age 8 to late adolescence. J Am Acad Child Adolesc Psychiatry 2006;45:578-86

22. Mannuza S, Klein RG, Bessler A, et al. Educational and occupational outcome of hyperactive boys grown up. J Am Acad Child Adolesc Psychiatry 1997;36:1222-7

23. Rappley MD. Attention deficit-hyperactivity disorder. N Engl J Med 2005;352:165-73

24. Arnold LE. Alternative treatments for adults with attention-deficit hyperactivity disorder (ADHD). Ann NY Acad Sci 2001;931:310341

25. National Institute for Clinical Excellence [NICE]. Health Technology Appraisal: Methylphenidate, atomoxetine and dexamfetamine for attention deficit hyperactivity disorder (ADHD) in children and adolescents including review of existing guidance number 13 (Guidance on the Use of Methylphenidate [Ritalin, Equasym] for Attention Deficit/Hyperactivity Disorder [ADHD] in childhood) – Scope. London: NICE, August 2003

26. National Institute for Health and Clinical Excellence [NICE]. Final Appraisal Determination: Methylphenidate, atomoxetine and dexamfetamine for attention deficit hyperactivity disorder (ADHD) in children and adolescents. London: NICE, May 2005

27. King S, Griffin S, Hodges Z, et al. A systematic review of the clinical and cost-effectiveness of methylphenidate hydrochloride, dexamfetamine sulphate and atomoxetine for attention deficit hyperactivity disorder (ADHD) in children and adolescents (commercial in confidence information removed). York: December 2004

28. Scottish Medicines Consortium. Atomoxetine capsules 10mg to 60mg (Strattera). No. 153/05. February 04, 2005, published online March 07, 2005, at: http://www.scottishmedicines.org.uk/medicines/default.asp. Last access February 15, 2006

29. Scottish Medicines Consortium. Atomoxetine capsules 10mg to 60mg (Strattera). No. 153/05. June 10, 2005, published online July 11, 2005, at: http://www.scottishmedicines.org.uk/medicines/default.asp. Last access February 15, 2006

30. Australian Government, Department of Health and Ageing. November 2005 PBAC Outcomes – Subsequent decisions not to recommend. Available online at: http://www.health.gov.au/internet/wcms/publishing.nsf/Content/pbacrec-pbacrecnov05-subsequent_rejections. November 2005. Last accessed February 15, 2006

31. Australian Government, Department of Health and Ageing. Atomoxetine hydrochloride, capsules, 10mg, 18mg, 25mg, 40mg and 60mg, StratteraR, November 2005. Available online at: http://www.health.gov.au/internet/wcms/publishing.nsf/Content/pbac-psd-atomoxetine-nov05. November 2005. Last accessed September 01, 2006

32. Banaschewski T, Coghill D, Santosh P, et al. Long-acting medications for the hyperkinetic disorders: a systematic review and European treatment guideline. Eur Child Adolesc Psychiatry 2006;May 5 (epub ahead of print)

33. Blades CA, Culyer AJ, Walker AM. Health service efficiency: appraising the appraisers: a critical review of economic appraisal in practice. Social Sci Med 1987;25:461-72

34. Schlander M. Commentary: Has NICE got it right? Curr Med Res Opin, 23: in press

35. National Institute for Clinical Excellence (NICE). Guide to the Technology Appraisal Process (reference N0514). London: NICE, April 2004

36. National Institute for Clinical Excellence (NICE). Guide to the Methods of Technology Appraisal (reference N0515). London: NICE, April 2004

37. Byford S, Palmer S. Common errors and controversies in pharmacoeconomic analyses. Pharmacoeconomics 1998;13:659-66

38. Neumann PJ, Stone PW, Chapman RH, et al. The quality of reporting in published cost-utility analyses, 1976-1997. Ann Intern Med 2000;132:964-72

39. Jefferson T, Demicheli V, Vale L. Quality of systematic reviews of economic evaluations in health care. JAMA 2002;287:2809-12

40. Schlander M. Health Technology Assessments by the National Institute for Health and Clinical Excellence: A Qualitative Study. New York, NY: Springer, 2007

41. Daniels N, Sabin J. The ethics of accountability in managed care reform. Health Aff 1998;17:50-64

42. Daniels N, Sabin JE. Setting Limits Fairly – Can We Learn to Share Medical Resources? Oxford: Oxford University Press, 2002

43. King S, Riemsma R, Hodges Z, et al. Technology Assessment Report for the HTA Programme: Methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder. Final version. London: NICE, June 2004. (Published online October 12, 2004: http://wwwnice.org.uk/page.aspx?o=adhd)

44. Dundar Y, Dodd S, Dickson R, et al. Comparison of conference abstracts and presentations with full-text articles in the health technology assessments of rapidly evolving technologies. Health Technol Assess 2006;10

45. King S, Griffin S, Hodges Z, et al. A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents. Health Technol Assess 2006;10

46. Newcorn JH, Owens JA, Jasinski DR, et al. Results from recently completed comparator studies with atomoxetine and methylphenidate. 51st Annual Meeting of the American Academy of Child & Adolescent Psychiatry (AACAP), Washington, DC: Symposium 20, October 21, 2004

47. National Institute for Health and Clinical Excellence (NICE): Appraisal of methylphenidate, atomoxetine and dexamfetamine for attention deficit hyperactivity disorder in children and adolescents: Appeal Panel Decision. London: NICE, December 8, 2005. Source: www.nice.org.uk/page.aspx?o=283566. Accessed December 20, 2005

48. National Institute for Health and Clinical Excellence [NICE]. Draft Scope: Attention deficit hyperactivity disorder: identification and management of ADHD in children, young people and adults. London: NICE, January 31, 2006. Source: www.nice.org.uk/page.aspx?o=290880. Accessed February 12, 2006

49. National Institute for Health and Clinical Excellence [NICE]. Final Scope: Attention deficit hyperactivity disorder: identification and management of ADHD in children, young people and adults. London: NICE, August 8, 2006. Source: www.nice.org.uk/page.aspx?o=351276. Accessed September 30, 2006

50. Taylor E, Sergeant J, Doepfner M, et al. Clinical guidelines for hyperkinetic disorder. Eur Child Adolesc Psychiatry 1998;7:184-200

51. Tripp G, Luk SL, Schaughency, EA, Singh R. DSM-IV and ICD-10: a comparison of the correlates of ADHD and hyperkinetic disorder. J Am Acad Child Adolesc Psychiatry 1999;38:156-64

52. Wolraich ML, Hannah JN, Baumgaertel A, Feurer ID. Examination of DSM-IV criteria for attention deficit/hyperactivity disorder in a county-wide sample. J Dev Behav Pediatr 1998;19:162-8

53. Taylor E, Doepfner M, Sergeant J, et al. European guidelines for hyperkinetic disorder – first upgrade. Eur Child Adoles Psychiatry 2004;13 (Suppl 1):7-30

54. Taylor E. Hyperkinetic disorders. In Gillberg C, Harrington R, Steinhausen H-C, eds. A Clinician's Handbook of Child and Adolescent Psychiatry. Cambridge: Cambridge University Press, 2006:489-521

55. Collett BR, Ohan JL, Myers KM. Ten-year review of rating scales. V. Scales assessing attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 2003;42:1015-37

56. American Psychiatric Association [APA]. Handbook of Psychiatric Measures. Washington, DC: American Psychiatric Association, 2000

57. Schachar R, Jadad AR, Gauld M, et al. Attention-deficit hyperactivity disorder: critical appraisal of extended treatment studies. Can J Psychiatry 2002;47:337-48

58. Jadad AR, Boyle M, Cunningham C, et al. Treatment of Attention-Deficit/Hyperactivity Disorder. Evidence Report / Technology Assessment No 11 (prepared by McMaster University under contract no 290-97-0017). AHRQ Publication No 00-E005. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ), November 1999

59. Danckaerts M, Heptinstall E, Chadwick O, Taylor E. Self-report of attention deficit hyperactivity disorder in adolescents. Psychopathology 1999;32:81-92

60. Conners CK. Conners' Rating Scales – Revised (CRS-R). In: American Psychiatric Association, Handbook of Psychiatric Measures. Washington, DC: American Psychiatric Association, 2000:329-32

61. Miller A, Lee SK, Raina P, et al. A review of therapies for attention-deficit/hyperactivity disorder. Ottawa, ON: Canadian Coordinating Office for Health Technology Assessment (CCOHTA), 1998

62. Klassen A, Miller A, Raina P, et al. Attention-deficit hyperactivity disorder in children and youth: a quantitative systematic review of the efficacy of different management strategies. Can J Psychiatry 1999;44:1007-16

63. MTA Cooperative Group. A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. Arch Gen Psychiatry 1999;56:1073-86

64. MTA Cooperative Group. Moderators and mediators of treatment response for children with attention-deficit/hyperactivity disorder: the multimodal treatment study of children with attention-deficit/hyperactivity disorder. Arch Gen Psychiatry 1999;56:1088-96

65. Biederman J, Quinn D, Weiss M, et al. Efficacy and safety of Ritalin LA, a new, once daily, extended-release dosage form of methylphenidate, in children with attention deficit hyperactivity disorder. Paediatr Drugs 2003;5:833-41

66. Newcorn J, Kratochvil CJ, Allen AJ, et al. Atomoxetine and OROS methylphenidate for the treatment of ADHD: acute results and methodological issues. Poster presentation at 45th Annual Meeting of the New Clinical Drug Evaluation Unit (NCDEU) of the National Institute of Mental Health (NIMH), Boca Raton, FL, June 6-9, 2005, Book of Abstracts, p. 188

67. Michelson D, Faries D, Wernicke J, et al. Atomoxetine in the treatment of children and adolescents with attention-deficit/hyperactivity disorder: a randomized, placebo-controlled, dose response study. Pediatrics 2001;1008: e83/1-9

68. Hunter JE, Schmidt FL. Dichotomization of continuous variables: the implications for meta-analysis. J Appl Psychol 1990;75:334-49

69. Guy W. ECDEU Assessment Manual for Psychopharmacology – Revised (DHEW Publ No ADM 76-338). Rockville, MD: Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, 1976:218-22

70. Beneke M, Rasmus W. 'Clinical Global Impressions' (ECDEU): some critical comments. Psychopharmacology 1992;25:171-6

71. Dahlke F, Lohaus A, Gutzmann H. Reliability and clinical concepts underlying global judgments in dementia: implications for clinical research. Psychopharmacol Bull 1992;28:425-32

72. Guy W. Clinical Global Impressions (CGI) Scale. In: American Psychiatric Association, Handbook of Psychiatric Measures. Washington, DC: American Psychiatric Association, 2000:100-2

73. Coghill D, Spencer Q, Barton J, et al. Measuring quality of life in children with attention-deficit/hyperactivity disorder in the UK. 16th World Congress of the International Association for Child and Adolescent Psychiatry and Allied Professions (IACAPAP). Book of Abstracts. Darmstadt: Steinkopff-Verlag, 2004:327

74. Secnik K, Cottrell S, Matza LS, et al. Assessment of health state utilities for attention-deficit/hyperactivity disorder in children using parent-based standard gamble scores. Value Health 2004;7:236

75. Sharp WS, Alter JM, Marsh WL, et al. ADHD in girls: clinical comparability of a research sample. J Am Acad Child Adolesc Psychiatry 1999;38:40-7

76. Arnold LE. Sex differences in ADHD: conference summary. J Abnorm Child Psychol 1996;24:555-69

77. Biederman J, Kwon A, Aleardi M, et al. Absence of gender effects on attention deficit hyperactivity disorder: findings in nonreferred subjects. Am J Psychiatry 2005;162:1083-9

78. Kemner JE, Starr HL, Brown DL, et al. Greater improvement and response rates with OROS MPH vs atomoxetine in children with ADHD. Presentation at the XXIVth Congress of the Collegium Internationale Neuro-Psychopharmacologicum, Paris, France, June 20-24, 2004

79. Kemner JE, Starr HL, Ciccone PE, et al. Outcomes of OROS methylphenidate compared with atomoxetine in children with ADHD: a multicenter, randomized prospective study. Advan Ther 2005;22:498-512

80. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. J Chronic Dis 1967;20:637-48

81. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on good research practices – modeling studies. Value Health 2003;6:9-17

82. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes, 2nd edn. Oxford: Oxford University Press,1997

83. Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the Economic Evaluation of Health Care Programmes, 3rd edn. Oxford: Oxford University Press, 2005

84. March JS, Silva SG, Compton S, et al. The case for practical clinical trials in psychiatry. Am J Psychiatry 2005;162:836-46

85. Ramsey S, Willke R, Briggs A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA task force report. Value Health 2005;8:521-33

86. Hughes DA, Bagust A, Haycox A, Walley T. Accounting for noncompliance in pharmacoeconomic evaluations. Pharmacoeconomics 2001;19:1185-97

87. Biederman J, Arnsten AFT, Faraone SV, et al. New developments in the treatment of ADHD. J Clin Psychiatry 2006;67:148-59

88. Weiss M, Gadow K, Wasdell MB. Effectiveness outcomes in attention-deficit/hyperactivity disorder. J Clin Psychiatry 2006;67 (suppl 8):38-45

89. Meredith PA. Achieving and assessing therapeutic coverage. In Métry J-M, Meyer UA, eds. Drug Regimen Compliance: Issues in Clinical Trials and Patient Management. Chichester: John Wiley, 1999:41-60

90. Swanson J. Compliance with stimulants for attention-deficit / hyperactivity disorder. Issues and approaches for improvement. CNS Drugs 2003;17:117-31

91. Wilens T. Subtypes of ADHD at risk for substance abuse. Presentation at the 157th Annual Meeting of the American Psychiatric Association; New York, NY: May 1-6, 2004

92. Graff Low K, Gendaszek AE. Illicit use of psychostimulants among college students: a preliminary study. Psychol Health Med 2002;7:283-7

93. Claxton A, Cramer J, Pierce C. A systematic review of the associations between dose regimens and medication compliance. Clin Ther 2001;23:1296-310

94. Métry J-M. Measuring compliance in clinical trials and ambulatory care. In: Métry J-M, Meyer UA, eds. Drug Regimen Compliance: Issues in Clinical Trials and Patient Management. Chichester: John Wiley, 1999:1-21

95. Peck C. Non-compliance and clinical trials: regulatory perspectives. In: Métry J-M, Meyer UA, eds. Drug Regimen Compliance. Issues in Clinical Trials and Patient Management. Chichester: John Wiley, 1999

96. Swanson JM, Kinsbourne M, Roberts W, Zucker K. A time-response analysis of the effect of stimulant medication on the learning ability of children referred for hyperactivity. Pediatrics 1978;61:21-9

97. Greenhill LL. Pharmacologic treatment of attention deficit hyperactivity disorder. Psychiatr Clin North Am 1992;15: 1-27

98. Kimko HC, Cross JT, Abernethy DR. Pharmacokinetics and clinical effectiveness of methylphenidate. Clin Pharmacokinet 1999;37:457-70

99. Greenhill LL, Perel JM, Rudolf G, et al. Correlations between motor persistence and plasma levels of methylphenidate-treated boys with ADHD. Int J Neuropsychopharmacology 2001;4:207-15

100. Pelham WE, Burrows-MacLean L, Gnagy EM, et al. Once-a-day Concerta methylphenidate versus t.i.d. methylphenidate in natural settings. Pediatrics 2000;121:126-37

101. Weinstein AG. Clinical management strategies to maintain compliance in asthmatic children. Ann Allergy Asthma Immunol 1995;74:304-10

102. American Academy of Child & Adolescent Psychiatry (AACAP). Practice parameter for the use of stimulant medications in the treatment of children, adolescents, and adults. J Am Acad Child Adolesc Psychiatry 2002;41 (2 Suppl):26-49S

103. Coghill D. Current issues in child and adolescent psychopharmacology. Part 1: Attention-deficit hyperactivity and affective disorders. Advan Psychiatr Treat 2003;9:86-94

104. Steinhoff K. Attention-deficit/hyperactivity disorder: medication treatment-dosing and duration of action. Am J Manage Care 2004;10: S99-106

105. Hack S, Chow B. Pediatric psychotropic medication compliance: a literature review and research-based suggestions for improving treatment compliance. J Child Adolesc Psychopharmacol 2001;11:59-67

106. Métry J-M. Measuring compliance in clinical trials and ambulatory care. In Métry J-M, Meyer UA, eds. Drug Regimen Compliance: Issues in Clinical Trials and Patient Management. Chichester: John Wiley, 1999:1-21

107. Miller AR, Lalonde CE, McGrail KM. Children's persistence with methylphenidate therapy: a population-based study. Can J Psychiatry 2004;49:761-8

108. Hwang P, Cosby A. Laberge ME. Compliance with three-times daily methylphenidate in children with attention-deficit/hyperactivity disorder. Value Health 2003;6:273

109. Lage M, Hwang P. Methylphenidate formulation is associated with accident/injury rate in children with ADHD. Value Health 2003;6:688

110. Lage M, Hwang P. Effect of methylphenidate formulation for attention deficit hyperactivity disorder on patterns and outcomes of treatment. J Child Adolesc Psychopharmacol 2004;14:575-81

111. Kemner JE, Lage MJ. Effect of methylphenidate formulation on treatment patterns and use of emergency room services. Am J Health Syst Pharm 2006;63:317-22

112. Kemner JE, Lage MJ. Impact of methylphenidate formulation on treatment patterns and hospitalizations: a retrospective analysis. Ann Gen Psychiatry 2006;5:1-8

113. Sanchez RJ, Crismon ML, Barner JC, et al. Assessment of adherence measures with different stimulants among children and adolescents. Pharmacotherapy 2005;25:909-17

114. Marcus SC, Wan GJ, Kemner JE, Olfson M. Continuity of methylphenidate treatment for attention-deficit/hyperactivity disorder. Arch Pediatr Adolesc Med 2005;159:572-8

115. Freemantle N, Blonde L, Bolinder B, et al. Real-world trials to answer real-world questions. Pharmacoeconomics 2005;23:747-54

116. Drummond MF, Knapp MRJ, Burns TP, et al. Issues in the design of studies for the economic evaluation of new atypical antipsychotics: the ESTO study. J Ment Health Policy Econ 1998;1:15-22

117. Baltussen R, Leidl R, Ament A. Real world designs in economic evaluation: bridging the gap between clinical research and policy-making. Pharmacoeconomics 1996;16:449-58

118. Revicki DA, Frank L. Pharmacoeconomic evaluations in the real world: effectiveness versus efficacy studies. Pharmacoeconomics 1999;15:423-34

119. Steele M, Riccardelli R, Binder C. Effectiveness of OROS-methylphenidate vs. usual care with immediate release methylphenidate in ADHD children. Presentation at the American Psychiatric Association (APA) annual meeting, New York, NY, May 1-6th, 2004

120. Steele M, Weiss M, Swanson J, et al. A randomized, controlled, effectiveness trial of OROS-methylphenidate compared to usual care with immediate-release-methylphenidate in attention-deficit-hyperactivity-disorder. Can J Clin Pharmacol 2006;13: e50-62

121. Swanson JM, Kraemer HC, Hinshaw SP, et al. Clinical relevance of the primary findings of the MTA: success rates based on severity of ADHD and ODD symptoms at the end of treatment. J Am Acad Child Adolesc Psychiatry 2001;40:168-79

122. Lord J, Paisley S. The clinical effectiveness and cost-effectiveness of methylphenidate for hyperactivity in childhood. London: National Institute for Clinical Excellence (NICE), Version 2, August 2000

123. Zupancic JAF, Miller A, Raina P, et al. Economic evaluation of pharmaceutical and psychological/behavioural therapies for attention-deficit/hyperactivity disorder. In: Miller A, Lee SK, Rain P, et al. (eds.) A Review of Therapies for Attention-Deficit/Hyperactivity Disorder. Ottawa, ON: Canadian Coordinating Office for Health Technology Assessment (CCOHTA), 1998

124. Jensen PS, Garcia JA, Glied S, et al. Cost-effectiveness of attention-deficit/hyperactivity disorder (ADHD) treatments: estimates based upon the MTA Study. 16th World Congress of the International Association for Child and Adolescent Psychiatry and Allied Professions (IACAPAP). Berlin, Germany, 2004. Book of Abstracts. Darmstadt: Steinkopff-Verlag 219

125. Schlander M, Jensen PS, Foster EM, et al. Kosteneffektivität alternativer Behandlungsstrategien der Aufmerksamkeitsdefizit/Hyperaktivitätsstörung (ADHS): Erste Daten aus der amerikanischen MTA-Studie. Monatsschr Kinderheilkd 2004;152: Suppl. 1

126. Annemans L, Ingham M. Estimating cost-effectiveness of Concerta OROS in attention-deficit/hyperactivity disorder (ADHD) – adapting the Canadian Coordinating Office for Health Technology Assessment's (CCOHTA) economic model of methylphenidate immediate release versus behavioural interventions from a parent's perspective. Value Health 2002;5:517

127. Schlander M. Cost-effectiveness of methylphenidate OROS for attention-deficit/hyperactivity disorder (ADHD): an evaluation from the perspective of the UK National Health Service (NHS). Value Health 2004;7:236

128. De Ridder A, De Graeve D. Estimating willingness to pay for drugs to treat ADHD – a contingent valuation study in students. Value Health 2002;5:462

129. Iskedjian M, Maturi B, Walker J, et al. Cost-effectiveness of atomoxetine in the treatment of attention deficit hyperactivity disorder in children and adolescents. Value Health 2003;3:275

130. De Civita M, Regier D, Alamgir AH, et al. Evaluating health-related quality-of-life studies in paediatric populations: some conceptual, methodological and developmental considerations and recent applications. Pharmacoeconomics 2005;23:659-85

131. Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. Pediatrics 2005;115: e600-14

132. Escobar R, Soutullo CA, Hervas A, et al. Worse quality of life for children with newly diagnosed attention-deficit/hyperactivity disorder, compared with asthmatic and healthy children. Pediatrics 2005;116: e364-9

133. Klassen A, Miller A, Fine S. Health-related quality of life in children and adolescents who have a diagnosis of attention-deficit/hyperactivity disorder. Pediatrics 2004;114:541-7

134. Sawyer MG, Whaites L, Rey JM, et al. Health-related quality of life of children and adolescents with mental disorders. J Am Acad Child Adolesc Psychiatry 2002;41:530-7

135. Fischer M, Barkley RA, Fletcher KE, Smallish L. The stability of dimensions of behaviour in ADHD and normal children over an 8-year follow-up. J Abnorm Child Psychol 1993;21:315-37

136. Barkley RA, Fischer M, Edelbrock CS, Smallish L. The adolescent outcome of hyperactive children diagnosed by research criteria. III. Mother-child interactions, family conflicts and maternal psychopathology. J Child Psychol Psychiatry 1991;32:233-55

137. Loeber R, Green SM, Lahey BB, Stouthamer-Loeber M. Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. J Abnorm Child Psychol 1991;19:75-95

138. Wells KC. Comprehensive versus matched psychosocial treatment in the MTA Study: conceptual and empirical issues. J Clin Child Psychol 2001;30:131-5

139. Wells KC, Pelham WE, Kotkin RA, et al. Psychosocial treatment strategies in the MTA Study: rationale, methods, and critical issues in design and implementation. J Abnorm Child Psychol 2000;28:483-505

140. Greenhill LL, Abikoff H, Arnold LE, et al. Medication treatment strategies in the MTA Study: relevance to clinicians and researchers. J Am Acad Child Adolesc Psychiatry 1999;34:1304-13

141. Greenhill LL, Swanson JM, Vitiello B, et al. Impairment and deportment responses to different methylphenidate doses in children with ADHD: the MTA titration trial. J Am Acad Child Adolesc Psychiatry 2001;40:180-7

142. Vitiello B, Severe JB, Greenhill LL, et al. Methylphenidate dosage for children with ADHD over time under controlled conditions: lessons from the MTA. J Am Acad Child Adolesc Psychiatry 2001;40:188-96

143. Owens EB, Hinshaw SP, Kraemer HC, et al. Which treatment for whom for ADHD? Moderators of treatment response in the MTA. J Consult Clin Psychol 2003;71:540-52

144. Jensen PS, Hinshaw SP, Kraemer HC, et al. ADHD comorbidity findings from the MTA Study: comparing comorbid subgroups. J Am Acad Child Adolesc Psychiatry 2001;40:147-58

145. Jensen PS, Garcia JA, Glied S, et al. Cost-effectiveness of ADHD treatments: findings from the multimodal treatment study of children with ADHD. Am J Psychiatry 2005;162:1628-36

146. Schlander M, Jensen PS, Foster EM, et al. Incremental costeffectiveness ratios of clinically proven treatments for attentiondeficit/hyperactivity disorder (ADHD): impact of diagnosticcriteria and comorbidity. 5th World Congress, InternationalHealth Economics Association (iHEA). Book of Abstracts. Barcelona, 2005:194-5

147. Foster M, Jensen PS, Schlander M, et al. Treatment for ADHD: Is More Complex Treatment Cost-Effective for More Complex Cases? Health Services Research 2007; 42: 165-182

148. Steinhoff K, Wigal T, Swanson J. Single daily dose ADHD medication effect size evaluation. Poster presentation, 50th Annual Meeting of the American Academy for Child and Adolescent Psychiatry, Miami, FL, October 22-27, 2003

149. Biederman J, Lopez FA, Boellner SW, Chandler MC. A randomized, double-blind, placebo-controlled, parallel group study of SLI381 (Adderall XR) in children with attention-deficit/hyperactivity disorder. Pediatrics 2002;110:258-66

150. Wolraich ML, Greenhill LL, Pelham WL, et al. Randomized, controlled trial of OROS methylphenidate once a day in children with attention-deficit/hyperactivity disorder. Pediatrics 2001;108:883-92

151. Michelson D, Allen AJ, Busner J, et al. Once-daily atomoxetine treatment for children and adolescents with attention deficit hyperactivity disorder: a randomized, placebo-controlled study. Am J Psychiatry 2002;159:1896-901

152. Faraone SV. Understanding the effect size of ADHD medications: implications for clinical care. Medscape 2003;8:1-7

153. Abikoff HG, Hechtman L, Klein RG, et al. Symptomatic improvement in children with ADHD treated with long-term methylphenidate and multimodal psychosocial treatment. J Am Acad Child Adolesc Psychiatry 2004;43:802-11

154. Klein RG, Abikoff HG, Hechtman L, Weiss G. Design and rationale of controlled study of long-term methylphenidate and multimodal psychosocial treatment in children with ADHD. J Am Acad Child Adolesc Psychiatry 2004;43:792-801

155. Kauffman RE, Smith-Right D, Reese CA, et al. Medication compliance in hyperactive children. Pediatr Pharmacol 1981;1:231-7

156. Firestone P. Factors associated with children's adherence to stimulant medication. Am J Orthopsychiatry 1982;252:447-57

157. Sleator EK, Ullmann RK, von Neumann A. How do hyperactive children feel about taking stimulants and will they tell the doctor? Clin Pediatr 1982;21:474-9

158. Brown RT, Borden KA, Clingerman SR. Adherence to methylphenidate therapy in a population: a preliminary investigation. Psychopharmacol Bull 1985;21:28-36

159. Brown RT, Borden KA, Wynne ME, et al. Compliance with pharmacologic and cognitive treatment for attention deficit disorder. J Child Adolesc Psychopharmacol 1987;26:521-6

160. Johnston C, Fine S. Methods of evaluating methylphenidate in children with attention deficit hyperactivity disorder: acceptability, satisfaction, and compliance. J Pediatr Psychol 1993;18:717-30

161. Greenhill LL, Findling RL, Swanson JM. A double-blind, placebo-controlled study of modified-release methylphenidate in children with attention-deficit/hyperactivity disorder. Pediatrics 2002;109: e39/1-7

162. Pliszka SR, Browne RG, Olvera RL, Wynne SK. A double-blind, placebo-controlled study of Adderall and methylphenidate in the treatment of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 2000;39:619-26

163. Faraone SV, Pliszka SR, Olvera RL, et al. Efficacy of Adderall and methylphenidate in attention deficit hyperactivity disorder: a reanalysis using drug-placebo and drug-drug response curve methodology. J Child Adolesc Psychopharmacol 2001;11:171-80

164. Klein RG, Abikoff H. Behavior therapy and methylphenidate in the treatment of children with ADHD. J Atten Disord 1997;2:89-114

165. Kelsey DK, Sumner CR, Casat CD, et al. Once-daily atomoxetine treatment for children with attention-deficit/hyperactivity disorder, including an assessment of evening and morning behavior: a double-blind, placebo-controlled trial. Pediatrics 2004;114: e1-8

166. Weiss M, Tannock R, Kratochvil C, et al. A randomized, placebo-controlled study of once-daily atomoxetine in the school setting in children with ADHD. J Am Acad Child Adolesc Psychiatry 2005;44:647-55

167. Spencer T, Heiligenstein JH, Biederman J, et al. Results from 2 proof-of-concept, placebo-controlled studies of atomoxetine in children with attention-deficit/hyperactivity disorder. J Clin Psychiatry 2002;63:1140-7

168. Elia J, Borcherding BG, Rapoport JL, Keysor CS. Methylphenidate and dextroamphetamine treatments of hyperactivity: Are there true nonresponders? Psychiatr Res 1991;36:141-55

169. Castellanos FX, Giedd JN, Elia J, et al. Controlled stimulant treatment of ADHD and comorbid Tourette's syndrome: effects of stimulant and dose. J Am Acad Child Adolesc Psychiatry 1997;36:589-96

170. Spilker B. Guide to Clinical Trials. New York, NY: Raven Press, 1991
171. NHS Centre for Reviews and Dissemination. CRD Report Number 24, Outcomes measurement in psychiatry: a critical review of outcomes measurement in psychiatric research and practice. York, NHS Centre for Reviews and Dissemination, July 2003
172. Nord E, Pinto JL, Richardson J, et al. Incorporating societal concern for fairness in numerical valuations of health programs. Health Econ 1999:8:25-39
173. Richardson J, McKie J. Empiricism, ethics and orthodox economic theory: what is the appropriate basis for decision-making in the health care sector? Social Sci Med 2005;60:265-75
174. Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. Health Econ 2005;14:197-208
175. Mortimer D. The value of thinly spread QALYs. Pharmacoeconomics 2006;24:845-53
176. Buxton M, Drummond MF, Van Hout BA, et al. Modelling in economic evaluation: an unavoidable fact of life. Health Econ 1997, 6:217-27
177. Porzsolt F, Kajnar H, Awa A, et al. Validity of original studies in health technology assessment reports: significance of standardized assessment and reporting. Int J Health Technol Assess Health Care 2005;21:410-13

Appendix

The following tables are available as electronic supplementary data (doi:10.1185/0300799078X260817) published with the online version of this article.

Table A1. *Synopsis of clinical studies selected by Assessment Group for primary ('base case') data synthesis and economic evaluation*

Table A2. *Synopsis of clinical studies added by assessment group for secondary ('extended') data synthesis and economic evaluation*

Table A3. *Some consistency issues related to the assessment report*