

INNOVATION AND VALUATION IN HEALTH CARE

Michael Schlander

Health Technology
Assessments by the
National Institute for Health
and Clinical Excellence

A Qualitative Study

 Springer

Summary

Results of Health Technology Assessments (HTAs) have become increasingly relevant to health care policy makers worldwide. The National Institute for Health and Clinical Excellence (NICE) in England is widely regarded as a role model for the implementation of HTAs, incorporating economic evaluation based on the logic of cost-effectiveness.

Beyond guidance on medical technologies to the National Health Service of England and Wales based on technology appraisals, NICE also issues clinical guidelines which are distinct from the HTAs in that their scope is usually broader, and in that development of the former is led by clinical experts, and the latter by economists.

The focus of this present report is the NICE appraisal process underlying its guidance concerning treatments for attention-deficit/ hyperactivity disorder (ADHD) in children and adolescents, issued in March 2006.

ADHD is of particular interest because it is associated with a number of descriptors that characterize a complex clinical decision problem.

Specifically:

- Prevalence estimates of ADHD vary depending on the population studied and diagnostic criteria used.
- ADHD is commonly associated with coexisting conditions, which include psychiatric comorbidities (such as “externalizing” symptomatology related to oppositional defiant disorder and conduct disorder, and “internalizing” symptomatology related to anxiety and depression) as well as a range of other psychiatric, neurological, and somatic disorders.
- International differences arise when different diagnostic criteria are used (e.g., DSM-IV in North America, and ICD-10 criteria in Europe) or when different standards of care and therapeutic preferences influence study designs, and this potentially confounds the interpretation of the results of studies across geographic boundaries.
- The diagnosis prevalence of ADHD appears to be increasing coincidentally with increased awareness of this condition and the increased use of stimulants in many countries, which in turn has led to controversy about the putative overuse of these interventions in this population.

- The variety of assessment instruments used to measure clinical outcomes in ADHD, the inherent variability in calculating health-related quality of life (HRQoL) outcomes, alongside the emergence of new treatment options, many of which are associated with higher unit costs than earlier options, collectively exacerbate the difficulties of conducting HTAs of interventions in this condition.

The objective of the present report is to explore how NICE appraisal processes can accommodate these clinical complexities. To this end, a qualitative case study was done 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. The data for this study consisted of all relevant technical documents produced by NICE (including meeting minutes and announcements) that were made publicly available (www.nice.org.uk). All key steps of the appraisal process were identified. In addition, a comprehensive review of the literature on ADHD treatment strategies was conducted.

The NICE appraisal process consists of three stages – “scoping”, “assessment”, and “appraisal”, each of them offering defined opportunities for stakeholders to provide input into the evaluation process. A fourth component of the appraisal process is “appeal” by consultees when there are predefined grounds for doing so.

NICE had reported a first appraisal of the use of interventions for hyperkinetic disorder (ADHD per ICD-10) in October 2000. It recommended the use of methylphenidate as part of a comprehensive treatment program for severe ADHD.

NICE subsequently reviewed the evidence and in 2005 concluded that:

- Where drug treatment is considered appropriate, methylphenidate, atomoxetine, and dexamphetamine are recommended within their licensed indications.
- There are no significant differences between individual drugs in terms of efficacy or side effects – a conclusion derived as a consequence of paucity of evidence used for assessment.
- Given the limited data used to inform response and withdrawal rates, it is not possible to distinguish between the different strategies on the grounds of cost-effectiveness.

- If there is a choice of more than one appropriate drug, the product with the lowest cost should be prescribed.

In the underlying assessment, the economic model was, in the absence of identified effectiveness differences, driven by drug acquisition costs, and a treatment strategy had been recommended as “clearly optimal” that consisted of 1st line dexamphetamine sulphate, 2nd line methylphenidate hydrochloride (immediate-release formulations), and 3rd line atomoxetine hydrochloride.

The Final Appraisal Determination further stated that the decision about choice of intervention should be based on:

- The presence of comorbid conditions (e.g., tic disorders, Tourette’s syndrome, epilepsy).
- The adverse event profile.
- Compliance issues (e.g., the need to administer a mid-day dose at school, and its associated implications).
- The individual preferences of the patient and/or parent/guardian.

Final guidance was published in March 2006, after an appeal had been dismissed, and reflected the Final Appraisal Determination. Clinical guidelines are in preparation.

Key aspects of the critique include the following:

Critique 1, Scoping.

- Despite the documented importance of psychosocial interventions in ADHD, this remained beyond the scope of the NICE appraisal. This omission, therefore, precludes the evaluation of such interventions alongside drug therapy. Existing evidence was not used to its full potential. In contrast, the scope for clinical guideline development does encompass psychosocial treatment.

- Although part of the scope, the assessment failed to address the potential impact of diagnostic criteria and coexisting conditions on the clinical and cost-effectiveness of treatment strategies assessed.

Critique 2, data selection by the Assessment Group.

- The effectiveness review of the HTA focused on hyperactivity measures at the expense of other ADHD-defining core symptoms of inattention and impulsivity.
- One third of the studies selected for effectiveness review were <3 weeks treatment duration, in violation of inclusion criteria defined via the assessment protocol, which were introduced to ensure sufficient time to evaluate the impact of treatment on indicators of social adjustment.
- A number of high-quality, double-blind trials were either discounted or overlooked by the Assessment Group because they did not fit alongside a predetermined model.
- For the economic model, in order to enable QALY calculation based on clinical response rates, the most widely used ADHD-specific outcomes instrument – the Conners scales – was not used by the Assessment Group. Instead, responders were defined by improvement of Clinical Global Impression (CGI) subscale scores, which are – besides their problematic psychometric properties – not wholly appropriate to provide normative information independent from baseline.
- The choice of the CGI-I (improvement) subscale as the primary outcome measure for cost-effectiveness evaluations by the Assessment Group substantially reduced the number of studies included in the economic analysis. Therefore, for one treatment modality (which was subsequently recommended for first-line treatment) a small-scale short-term study had to be added that had been excluded before for quality concerns.
- As ADHD is a chronic disorder, long-term treatment- and cost-effectiveness considerations are important. Yet, the Assessment Group evaluated studies of 3 to 8 weeks treatment duration, despite the availability of at least 15 randomized trials of ≥ 12 weeks treatment duration (although not all of these contained all of the elements selected for extraction by the Assessment Group, such as CGI-I data).

Critique 3, efficacy, effectiveness, and the role of treatment compliance.

- The Assessment Group did not address the wide-ranging issues surrounding the importance of the distinction between clinical efficacy and effectiveness.
- The assumptions made by the Assessment Group about the measurement and impact of noncompliance in ADHD remained a major issue throughout the appraisal. In effect, any clinical benefit potentially resulting from improved compliance was “assumed away” by the Assessment Group, without reference to the extensive literature on the subject.
- Artificially enhanced compliance in randomized clinical trials (RCTs) is an important confounder to their external validity in a “real world” setting. Reduced or non-compliance to stimulants in ADHD would likely manifest as a rapid return to symptoms. However, the potential implications of this were not considered in the NICE appraisal.
- This is important because data indicate that the majority of children with ADHD miss doses and/or do not refill prescriptions. Moreover, due to its PK/PD properties, methylphenidate is “unforgiving” of missed doses, and multiple dosing of immediate-release formulations throughout the day is required to maintain effectiveness. In addition, mid-day dosing may have social as well as compliance implications.
- Real-world evidence indicates that modified-release formulations of methylphenidate are associated with high response rates – perhaps as a result of reducing the noncompliance risk – a factor that was not adequately considered in the NICE appraisal.

Critique 4, data synthesis across endpoints and studies.

- Evidence remaining after selection of data for the primary economic analysis was insufficient to assess the relative value of alternative treatment approaches.
- In an attempt to overcome this shortcoming, the Assessment Group synthesized response rates across different effectiveness measures, and used the statistical mixed treatment comparison (MTC) technique to facilitate the integration of both direct and indirect evidence.

- To broaden the dataset available for analysis, the Assessment Group imported data from additional trials that reported different outcome measures (derived from clinical global impressions and narrow-band symptom scales). This raises important questions about the potential for the use of heterogeneous outcome measures to confound the conclusions.
- Although the use of meta-analyses is a well-accepted approach, the validity of this approach is a function of the internal validity and similarity of the trials to be included and assumes that relative treatment effects will be the same across trials. In the present ADHD assessment, this approach necessarily concealed effects of enhanced compliance associated with long-acting medications, including but not limited to modified-release formulations of methylphenidate, in real-world settings, thus introducing a bias against this group of medications.
- In the NICE appraisal, there were multiple sources of potential bias, relating to the small number of studies selected, wide variations in patient numbers between selected studies, short observation periods, heterogeneity of studies (including heterogeneous populations, study designs and endpoints), and compliance issues.

Critique 5, economic model.

- Symptom scales used for the assessment of ADHD do not usually qualify as instruments to measure HRQoL outcomes, contrary to their interpretation in the NICE Assessment Report.
- Deviating from the search strategy defined in the assessment protocol, several key publications were omitted, which may have impacted on the overall outcome of the appraisal.
- The structure of the model resulted in double-counting of nonresponders, which affected treatment options differently and was a potential source of distortion and bias.
- Compliance issues were effectively excluded from consideration in the economic model; three to eight (or twelve) weeks treatment duration in controlled trials were assumed to capture long-term treatment persistence.

- Although the economic model was extended to a time horizon of 12 years, clinically relevant long-term sequelae associated with ADHD were not addressed.

Critique 6, appraisal and appeal process.

- The Appraisal Consultation Document noted the ADHD core signs of inattention, hyperactivity, and impulsiveness, the difference between ICD-10 and DSM-IV definitions, and the potential influence of comorbidity on therapeutic outcomes in ADHD, although the Assessment Report had failed to adequately address those.
- The Appraisal Committee found that methodological flaws in some studies limited their persuasive value. However, the “flaw” of being open-label was an essential design component of a pragmatic real world study considered, increasing its external validity.
- An appeal was lodged on the basis of the omission of a key study from the assessment process, which might have influenced the Final Appraisal Determination. However, the appeal was dismissed.

Against the background of these observations, the conclusions of NICE are contrasted with insights from clinical long-term studies, from disease-specific effectiveness measures, and from other HTAs concerned with ADHD treatment strategies.

Specifically:

A number of long-term trials in ADHD have become available that may enable differentiation between treatment approaches, also by diagnostic criteria, by comorbidity, and by intensity of treatment. Retrospective database analyses lend support to the issue of treatment persistence and indicate a differential impact of the products assessed. Furthermore, evidence is growing that nonstimulant interventions (e.g., atomoxetine), which are more expensive, are not more (or less) effective than stimulant drugs (e.g., methylphenidate) might have implications for NICE guidance as well. These conclusions receive further support from concordant results of HTAs of ADHD treatment strategies in other jurisdictions.

It is therefore concluded that the NICE appraisal and guidance does not adequately reflect current knowledge of ADHD and its treatment. This however is more than an academic issue because it may have clinical practice implications.

With reference to the case analysis, four distinct domains of problems underlying the anomalies observed are suggested.

– *Separation of clinical and economic expertise.*

The appraisal was driven by economists without sufficient integration of clinical perspectives and expertise. This may have contributed, inter alia, to not taking into account the impact on the appraisal of different measurement instruments for ADHD, the role of treatment compliance in the clinical effectiveness of ADHD treatments, and the importance of adequate treatment duration.

– *High level of standardization.*

The drive to establish consistency and transparency across the technology appraisal process has reduced the rich clinical evidence base to a few short-term studies reporting clinical global improvements on a subscale that might be less than optimal, potentially resulting in bias and misleading results. An alternative approach – that is, to seek a solution specific to the condition and the available data, rather than define a problem to fit a common framework – may be a more pragmatic approach.

– *Technical quality of assessment.*

Multiple shortcomings of a technical nature – such as a departure from search criteria specified in the assessment protocol, deviations from NICE reference case guidance, and a range of further anomalies and inconsistencies – may highlight an apparent shortfall in expected quality assurance systems for the technology assessment process, in addition to insufficient integration of clinical perspectives.

– *Process-related issues.*

Within the highly structured NICE process, transparency is limited when commercial-in-confidence data are used. This may impede effective stakeholder participation. Transparency is also limited with regard to uninformative Appraisal Committee meeting minutes and to economic models used by Assessment Groups. This is a potentially serious constraint since technology appraisals rely heavily on Assessment Reports, and transparency is broadly considered a key feature of model quality.

This notwithstanding, the predictability, inclusiveness, and overall publicity of the NICE appraisal process should be acknowledged.

Referring to the “accountability for reasonableness” (A4R) framework proposed by Daniels and Sabin, however, NICE falls short of expectations despite an official commitment to adhere to the framework. Transparency is incomplete, appeals are restricted and do not allow to reopen debate, and the extent that the conditions of relevance and enforcement are met may be subject to debate.

Suggested implications for international policy makers include consideration of:

- The objectives of health care provision, since the underlying social value judgments made by NICE are not universally shared.
- The extent of reliance on QALYs as an outcome measure, as the narrow analytical focus of NICE was identified as a prime reason for the selective use of the clinical evidence, which simultaneously caused exclusion of a wider evidence base that was inconsistent with QALY-based outcomes.
- Flexibility in use of analytic approaches, as greater flexibility compared with the NICE reference case might drive alternative evaluation techniques and provide increasingly robust guidelines.
- The technology appraisal process. Although relatively transparent, the NICE approach does not fully meet A4R criteria. Also NICE transparency is impeded by commercial-in-confidence data (submitted by manufacturers of technologies assessed) and intellectual property rights (concerning models developed by review teams) that restrict third-party participation in technology appraisals.
- The timing of technology appraisals. A fundamental paradox of cost-effectiveness evaluations is that early data are needed for policy impact, while reliable data require practical experience. This has been cited as a reason why modeling in economic evaluation is “an unavoidable fact of life”. At the same time, this should encourage to strive for more consistent use of information from studies not yet published in peer-reviewed journals, such as abstracts and conference presentations.
- The use of multidisciplinary assessment teams. The present case study of NICE’s ADHD appraisal highlights some of the constraints that may arise when economists

and clinicians work independently. The case supports the conception that greater integration of these disciplines at all stages of the appraisal process would more likely achieve the goals of each.

- Quality assurance. Standardization is not synonymous with quality consistency of HTAs. Apparently, additional precautions are warranted to achieve greater transparency in the quality assurance of the NICE process. Specific processes should be implemented to ensure high quality of evidence synthesis and economic models.
- Implementation, which may be achieved best when guidance is concordant with clinical needs and expectations, instead of defining separate clinical and decision-making perspectives, as was the case in the ADHD assessment.

Based upon the case analysis, possible ways forward are briefly discussed, highlighting different international starting points, key ethical aspects related to the objectives of health care provision, institutional context, and a research agenda to take matters further.

In conclusion, international health care policy makers contemplating to adopt NICE-like approaches appear well advised to consider both strengths and limitations of the NICE approach, in addition to the specific value judgments underlying NICE technology appraisals, which they may or may not share.