doi:10.1185/030079908X260817
Electronic supplementary data published in conjunction with Schlander M. Is NICE infallible? A qualitative study of its assessment of treatments for attention-deficit/hyperactivity disorder (ADHD) *Curr Med Res Opin* 2008;24:515-535

All reference numbers refer to the reference list in the above article

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

Clinical study			Patients				Treatments				Clinical endpoints				Comments
Authors	Design	*n*	Gender (male)	Age (mean/ range)	Subtypes	Co-morbidity	Modality: Drug	Modality: Dose[a]	NDT	Duration	CGI-I [RR]	CGI-S [RR]	ADHD-RS [RR]	SNAP-IV [RR]	
Sharp *et al.*, 1999[75]	3× cross-over	32	Girls only	?	C: 100% (?)	?				3m	Yes	?	No	No	Excluded in AR from effectiveness review (for reason of 'inadequate data presentation'); no further data provided in AR; inclusion 'initially' based on DSM-IIIR, 'later' DSM-IV, combined type
		32	0	?	C: 100% (?)	?	MPH-IR b.i.d.	1.28 mg/kg/ dose	'recreation therapy activities'	3w	81%	?	—	—	
		32	0	?	C: 100% (?)	?	DEX b.i.d.	0.64 mg/kg/ dose		3w	84%	?	—	—	
		32	0	?	C: 100% (?)	?	Placebo	n.a.		3w	16%	?	—	—	
Greenhill *et al.*, 2002[161]	RCT, double-blind, PG (1:1), 32 sites	314	257 (82%)	9 (5–15)	?	None (?)			None (?)	3w	Yes	No	No	No	Primary endpoint: Conners' Teacher Global Index; study listed among MPH-ER medium dose group in AR
		155	128 (83%)	9 (6–15)	?	None (?)	MPH-MR08 o.a.d.	40.7 mg/d	None (?)	3w	81%	63%	—	—	
		159	129 (81%)	9 (5–14)	?	None (?)	Placebo	n.a.	None (?)	3w	50%	26%	—	—	
Kemner *et al.*, 2004[78,79] Note that 'Kemner *et al.*, 2004' is quoted in the AR as 'CIC'	RCT, open-label, PG (2:1), 'multiple sites'	1323	982 (74%)	8.9 ± 2.1 (6–12)	C: 75% I: 14% H: 12%	?			None (?)	3w	Yes	?	Yes	No	'CIC' (no data provided in AR); primary endpoint: ADHD-RS improvement (change in mean score): MPH-MR12 superior to ATX; adherence ≥92% in both groups
		850	630 (74%)	8.8 ± 2.0 (6–12)	C: 74% I: 13% H: 13%	?	MPH-MR12 o.a.d.	32.7 mg/d	None (?)	3w	69%	?	76%[b]	—	
		473	352 (74%)	9.2 ± 2.1 (6–12)	C: 76% I: 15% H: 10%	?	ATX o.a.d. or b.i.d.	36.7 mg/d	None (?)	3w	53%	?	63%[b]	—	

Table A1 *(continued)*

Clinical study		Patients					Treatments				Clinical endpoints				Comments
Authors	Design	*n*	Gender (male)	Age (mean/ range)	Subtypes	Co-morbidity	Modality			Duration	CGI-I [RR]	CGI-S [RR]	ADHD-RS [RR]	SNAP-IV [RR]	
							Drug	Dose[a]	NDT						
Steele *et al.*, 2004[119], 2006[120] Note that 'Steele *et al.*, 2004' is quoted in the AR as 'CIC'	RCT, open-label, "real-world" design, PG (1:1)	73	60 (82%)	9.1 ± 1.8 (6–12)	C: 79% I: 19% H: 1%	ODD: 38%; CD: 0%	MPH-IR t.i.d. (61% of patients)	33.2 mg/d	None (?)	8w	62%	'CIC' mean change available only	—	16%	'CIC' (no data provided in AR); primary endpoint: SNAP-IV (18/26 items, parent ratings); real-world effectiveness trial; MPH-MR12 superior to MPH-IR
		72	61 (85%)	9.0 ± 2.1 (6–12)	C: 79% I: 18% H: 3%	ODD: 43%; CD: 1%	MPH-MR12 o.a.d.	37.8 mg/d	None (?)	8w	85%[c]		—	44%	
Pliszka *et al.*, 2000[162]; cf. also Faraone *et al.*, 2001[163]	RCT, double-blind, PG (1:1:1)	58	?	8.1 ± 1.4	?	ODD, CD, anxiety			None (?)	3w	Yes	No	No	No	Primary endpoint: IOWA Conners' ratings
		20	?	?	?	ODD, CD, anxiety	MPH-IR o.a.d. – t.i.d.	25.2 mg/d	None (?)	3w	65%	—	—	—	
		20	?	?	?	ODD, CD, anxiety	MAS o.a.d or b.i.d.	12.5 mg/d	None (?)	3w	90%	—	—	—	
		18	?	?	?	ODD, CD, anxiety	Placebo	—	None (?)	3w	28%	—	—	—	
Klein and Abikoff, 1997[164]	RCT, double-blind, PG (1:1:1)	86	81 (94%)	7.8 ± 1.4 (6–12)	?	None[d] (?)				8–12w	Yes	No	No	No	Primary endpoints: CTRS, CPRS; multiple further assessments
		29		?	?	None[d] (?)	MPH-IR b.i.d. or t.i.d. (?)	1.55 mg/kg/ d	None	8w[e]	79%[e]	—	—	—	
		29		?	?	None[d] (?)	MPH-IR b.i.d. or t.i.d. (?)	1.48 mg/kg/ d	Yes	8w[e]	97%[e]	—	—	—	
		28		?	?	None[d] (?)	Placebo	—	Yes	8w[e]	50%[e]	—	—	—	

All patients defined by DSM-IV diagnostic criteria. Subtypes: C, combined; I, inattentive; H, hyperactive. AR, Assessment Report[27], MPH, methylphenidate, MPH-IR, methylphenidate immediate-release; MPH-MR08, MPH-MR12, methylphenidate modified-release with a duration of action of 8 or 12 hours, respectively; ATX, atomoxetine, DEX, dexamphetamine, MAS, mixed amphetamine salts, trade name Adderall (not available in Europe); PG, parallel-group; ODD, oppositional-defiant disorder; CD, conduct disorder; CIC, commercial-in-confidence; NDT, non-drug treatment; RR, response rate; o.a.d., one dose per day; b.i.d., divided in two doses per day; t.i.d., divided in three doses per day; mg/d, average dose in milligrams per day; mg/kg/d, average dose in milligrams per kilogram bodyweight and day; w, week(s); [a]mean doses at study end; [b]definition of 'response': reduction in mean ADHD-RS score ≥30%; [c]Table 6.2 of AR gives a figure of 83%; [d]according to inclusion criteria 'relatively free of anxiety, depression, and conduct disorder'; [e]psychiatrist CGI-I ratings after 8 weeks of treatment, not 12 weeks as implied in Table 4.26 of AR

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

Clinical study		Patients					Treatments				Clinical endpoints				Comments
Authors	Design	*n*	Gender (male)	Age (mean/ range)	Subtypes	Co-morbidity	Modality			Duration	CGI-I [RR]	CGI-S [RR]	ADHD-RS [RR]	SNAP-IV [RR]	
							Drug	Dose[a]	NDT						
Kelsey *et al.*, 2004[165]	RCT, double-blind, PG (2:1), 12 sites	197	139 (71%)	9.5 (6–12)	C: 69% I: 27% H: 4%	ODD: 35.0%; CD: 4.1%				8w	No	Yes	Yes	No	Primary endpoint: ADHD-RS total score
		133	94 (71%)	9.5 ± 1.8 (6–12)	C: 70% I: 26% H: 4%	ODD: 37.6%; CD: 5.3%	ATX o.a.d.	44.5 mg/d	None (?)	8w	—	27%	63%	—	
		64	45 (70%)	9.4 ± 1.8 (6–12)	C: 67% I: 30% H: 3%	ODD: 29.7%; CD: 1.6%	Placebo	—	None (?)	8w	—	5%	33%	—	
Michelson *et al.*, 2002[151]	RCT, double-blind, PG (1:1), 9 sites	171		? (6–16)	C: 58% I: 41% H: 2%	ODD 20%; few others			None (?)	6w	No	Yes	Yes	No	Primary endpoint: ADHD-RS total score; other scores including CTRS and parent ratings of behavior, besides CGI-S
		85	60 (71%)	? (6–16)	C: 55% I: 41% H: 4%	ODD: 18.8%; few others	ATX o.a.d.	Not reported (1.0–1.5 mg/kg/d)	None (?)	6w	—	29%	60%	—	
		86	60 (71%)	? (6–16)	C: 60% I: 40% H: 0%	ODD: 21.1%; few others	Placebo	—	None (?)	6w	—	10%	31%	—	
Weiss *et al.*, 2004[27], 2005[166] Note that 'Weiss *et al.*, 2004' is quoted in the AR as 'CIC'	RCT, double-blind, PG (2:1), 11 sites (USA, Canada, Puerto Rico)	153	123 (80%)	9.9 ± 1.3 (8–12)	C: 73% I: 27% H: 1%	ODD 33%; learning disorders (LD): 30%; others: <5%				7w	Yes	Yes	Yes	No	'CIC' (no data provided in AR); primary endpoint: ADHD-RS total score (by teachers); further endpoints including CPRS and Conners' Global Index
		101	83 (82%)	9.9 ± 1.4	C: 74% I: 26% H: 0%	ODD: 33%; LDs: 29%	ATX o.a.d.	1.33 mg/kg/d		7w	?	21%	69%	—	
		52	40 (77%)	9.9 ± 1.3	C69% I: 29% H: 2%	ODD: 35%; LDs: 31%	Placebo	n.a.		7w	?	10%	43%	—	

Table A2 *(continued)*

Clinical study		Patients					Treatments				Clinical endpoints				Comments
Authors	Design	*n*	Gender (male)	Age (mean/ range)	Subtypes	Co-morbidity	Modality			Duration	CGI-I [RR]	CGI-S [RR]	ADHD-RS [RR]	SNAP-IV [RR]	
							Drug	Dose[a]	NDT						
Spencer *et al.*, 2002[167] (I)[a]	RCT, double-blind, PG (1:1), multi-center	147 (20 MPH-treated – no data)	201/253 (79%)[a]	(7–13)[a]	C. 80%[a] I: 19%[a] H: 1%[a]	ODD: 38.7%; phobias: 11.5%; some others[a]			None (?)	9w	No	No	Yes	No	Primary endpoint: ADHD-RS total score; further endpoints including subscales, CPRS, CGI-ADHD-Severity; an MPH arm comprising 20 patients was included 'to validate study design in the event that ATX failed to separate from placebo'; no MPH results are reported.
		65	98/129 (76%)[a]	9.7 ± 1.6 (7–13)[a]	C: 80.6%[a] I: 19%[a] H: 1%[a]	ODD: 41.1%; phobias: 12.4%; some others	ATX b.i.d., t.i.d.	Not reported (≤2.0 mg/kg/d or ≤90 mg/d)	None (?)	9w	—	—	65%	—	
		62	103/124 (83%)[a]	10.0 ± 1.5 (7–13)[a]	C: 79%[a] I: 19%[a] H: 2%[a]	ODD: 36.3%; phobias: 10.5%; some others[a]	Placebo	—	None (?)	9w	—	—	24%	—	
		20	?	?	?	?	MPH	?	?	9w (?)	?	?	?	?	
Spencer *et al.*, 2002[167] (II)[a]	RCT, double-blind, PG (1:1), multi-center	126 (144 ? - not clear: MPH)	201/253 (79%)[a]	(7–13)[a]	C. 80%[a] I: 19%[a] H: 1%[a]	ODD: 38.7%; phobias: 11.5%; some others[a]			None (?)	12w	No	No	Yes	No	Primary endpoint: ADHD-RS total score; further endpoints including subscales, CPRS, CGI-ADHD-Severity
		64	98/129 (76%)[a]	9.7 ± 1.6 (7–13)[a]	C: 80.6%[a] I: 19%[a] H: 1%[a]	ODD: 41.1%; phobias: 12.4%; some others[a]	ATX b.i.d., t.i.d.	Not reported (≤2.0 mg/kg/d or ≤90 mg/d)	None (?)	12w	—	—	59%	—	
		62	103/124 (83%)[a]	10.0 ± 1.5 (7–13)[a]	C: 79%[a] I: 19%[a] H: 2%[a]	ODD: 36.3%; phobias: 10.5%; some others[a]	Placebo	—	None (?)	12w	—	—	44%	—	

Table A2 *(continued)*

Clinical study		Patients					Treatments				Clinical endpoints				Comments
Authors	Design	*n*	Gender (male)	Age (mean/ range)	Subtypes	Co-morbidity	Modality			Duration	CGI-I [RR]	CGI-S [RR]	ADHD-RS [RR]	SNAP-IV [RR]	
							Drug	Dose[a]	NDT						
Swanson *et al.*, 2001[121]; cf. also MTA Cooperative Group, 1999[63,64]	RCT, open-label, PG (1:1:1:1), 6 sites	579	465 (80%)	8.5 ± 0.8 (7– <9.9)		Int: 14.0%; Ext: 29.5%; Both: 24.7%			See below	14m (24m)	No	No	No	Yes	MTA study[b] A variety of endpoints was studied, including SNAP-IV, CTRS, CPRS, CIS SNAP-IV-based response rates were calculated using parent and teacher ratings, including all 26 items of the scale (Swanson *et al.*, 2001[121]) Results for NDT alone (behavioral treatment arm) 'omitted from evaluation' by the assessment group 'as not relevant to this review' (AR, p. 254)
		144	118 (82%)	8.6 ± 0.8 (7– <9.9		Int.: 13.9%; Ext: 27.8%; Both: 26.4%	Medication management: MPH-IR t.i.d.	75%: MPH-IR 37.7 mg/d	None	14m (24m)	—	—	—	56% (14m)	
		145	114 (79%)	8.4 ± 0.8 (7– <9.9		Int.: 13.1% Ext: 24.8%; Both: 25.5%	Combination treatment: MPH-IR t.i.d.	75%: MPH-IR 31.2 mg/d	Yes	14m (24m)	—	—	—	68% (14m)	
		146	119 (82%)	8.5 ± 0.8 (7– <9.9		Int.: 13.0%; Ext: 37.0%; Both: 21.1%	Community comparison	58%: MPH-IR 22.6 mg/d, mean 2.3 doses/d	Partly	14m (24m)	—	—	—	25% (14m)	
		144	114 (79%)	8.3 ± 0.8 (7– <9.9		Int.: 16.0%; Ext: 29.2%: Both: 25.0%	None (behavioral management)	—	Yes	14m (24m)	—	—	—	34% (14m)	
Quinn *et al.*, 2003 [cf. AR, p. 93, p. 146, p. 276][27]	? ('CIC')	?	?	?	?	?	?		?	?	?	?	?	?	'CIC' (no data provided in AR); source quoted: 'Celltech Integrated Clinical Study Report'
		?	?	?	?	?	MPH-IR	'High dose': > 30 mg/d	?	?	?	?	?	?	
		?	?	?	?	?	'MPH-ER'; presumably MPH-MR08	'Med. dose': 20–40 mg/d	?	?	?	?	?	?	
		?	?	?	?	?	Placebo	—	?	?	?	?	?	?	

Table A2 *(continued)*

Clinical study		Patients					Treatments				Clinical endpoints				Comments
Authors	Design	*n*	Gender (male)	Age (mean/ range)	Subtypes	Co-morbidity	Modality			Duration	CGI-I [RR]	CGI-S [RR]	ADHD-RS [RR]	SNAP-IV [RR]	
Elia *et al.*, 1991[168]; see also Castellanos *et al.*, 1997[169]	3× cross-over	48	48 (100%)	8.6 ± 1.7 (6–12)	DSM III No data on subtypes reported	ODD: 25.0%; CD: 20.8%; specific development disorders 22.9%			Yes	9w	'CGI results were presented in graph form only and could not be reproduced in a table.' (AR, p. 156)	No	No	This study used DSM-III diagnostic criteria and defined response as a score of 1, 2, or 3 on CGI-I; primary endpoint not specified
							DEX b.i.d.	≤1.3 mg/kg/d	Yes	3w		—	—	
							MPH-IR b.i.d.	≤2.5 mg/kg/d	Yes	3w		—	—	
							Placebo	—	Yes	3w		—	—	

For abbreviations used, cf. legend to Table A1; int.: internalizing comorbidity (anxiety, depression), ext.: externalizing comorbidity (CD, ODD). [a]Spencer *et al.* (2002)[167] reported on two randomized trials; some data are provided only for both trial populations combined; [b]for the NIMH MTA Study, cf. *Discussion*; note that ADHD-RS response definitions varied across trials: response definition in the study by Weiss *et al.* (2004[27], 2005[166]) was '20% reduction in the ADHD-RS-IV Teacher: Inv total score' (l.c., p. 650), whereas Elia *et al.* (1991[168]) used different diagnostic criteria; this study was listed as a 9-weeks-study in the AR though treatment modalities were tested for 3 weeks each. Response definitions varied also between studies for SNAP-IV scores[119,120,121,166,167]

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

Subject	Statements and rationales	Consistency issue
Search criteria	'Economic evaluations reported as conference proceedings or abstracts were excluded since the data may not be complete' (AR, p. 178)	Departure from assessment protocol[61], which had promised to include abstracts, conference proceedings, gray literature, …'
		Violations of predefined search strategy, e.g. overlooked RCTs and CEAs in the public domain
		The incomplete search did not prevent from claiming that 'the review highlighted a number of potential limitations in the existing literature … in particular … in estimating treatment effectiveness …, [which] may stem from a lack of available data' (AR, p. 266)
		Inclusion of (at least) one study in economic model that had been excluded from the effectiveness review and not listed in appendix
	'This review presents a comprehensive overview of existing *economic* evaluations of MPH, ATX and DEX for children and adolescents with ADHD' (AR, p. 266)	Cost-effectiveness analyses in the public domain were excluded from consideration owing to the illicit change of search criteria (cf. above)
Study inclusion criteria	Minimum study duration was chosen because 'the literature suggests that three weeks is the minimum duration for therapeutic trials' assessing 'the impact on the social adjustment of the child' (AR, pp. 44f.)	More than one third of studies included in the effectiveness review were short-term crossover studies with *treatment duration* of one week or less, and some of them had been conducted without washout phases between treatment periods
		There was no review of 'the literature' supporting the assertion; except for one reference to the DSM-IV diagnostic manual (AR, p. 45)
		If social adjustment of a child is the clinical outcome of interest, then (a) a clinical effect measure capturing functional impairment (which were discarded: AR, p. 46) would have been more appropriate than CGI ratings, which were used as a proxy for health-related quality-of-life (for instance: AR, pp. 16, 17, 46, 48), and (b) crossover designs will be problematic due to frequent violation of the requirement that 'a similar baseline condition must be present at the start of each of the treatment periods, … and there must not be any carryover (i.e., residual) effects (even psychological ones) after either treatment'[170]
Outcome measures	A 'plethora of instruments' was noted (AR, p. 178) and it was recognized that 'the choice of an outcome measure is a critical design issue' (AR, p. 178)	These observations were not followed up by an exploration of the extensive literature on this subject area[55,56,171]
Conners Rating Scales	Rejected for economic modeling…	… although representing the most widely used effect measure in ADHD research to date[55,57,58,61], and the only one enabling quantitative synthesis in previous reviews[57,61]

Table A3 *(continued)*

Subject	Statements and rationales	Consistency issue
	… on the basis of a critique of the 'implicit assumption' that a small gain in CTRS score for many children is assumed to be the same as the cost and desirability of achieving a large gain in CTRS score for few children' (AR, p. 186). Further it was argued that, 'if this measure is used, a gain of 1 point on the scale is valued the same, regardless of where you begin on that scale, so the relative value of different effect sizes is not readily interpretable' (AR, p. 224)	A remarkably similar critique has been put forward (and been supported by a large body of empirical evidence showing that identical QALY differences will not be attached equal social values across the scale[172–175]) against QALYs, the outcome measure used for economic evaluation following NICE reference case guidance[36]
	… on the basis of a critique of the 'implicit assumption … that efficacy is constant across baseline levels of ADHD severity. … However, the efficacy of stimulants [or medication in general] *may* depend on the quality and severity of symptoms' (AR, p. 186)	The advantage of the CGI-I scores selected for primary evaluation remains unclear, as this score consists of one item only, reading 'Rate total improvement … compared to the patient's condition on admission to the project'[69] – hardly constituting a measure independent from baseline severity or commanding interval-scale properties
QALY calculation	For utility estimates it was explicitly recognized that 'the validity of these measures depends on the content and style of the vignette used to describe each health state' (AR, p. 181)	Health state descriptions for utility measurement (responders and non-responders) did not match the CGI (as well as other criteria synthesized) criteria (cf. AR, pp. 359ff.)
	Health state utilities derived from a company submission were used for extended sensitivity analyses (AR, p. 235, pp. 240ff.), …	… although inconsistencies of these values had been identified (AR, p. 217), and inspection of health state descriptions (e.g., AR, pp. 359ff.) reveals 'double-counting' of side effects for MPH-IR and MPH-MR
	Utility data for the primary economic evaluation came from 'values obtained using a *standard gamble* technique from parents of children with ADHD, providing proxy ratings for their children' (AR, p. 235)	In fact, these utilities were derived from *EQ-5D questionnaires* completed by parents or caregivers[73]
	Utility 'values *obtained directly from patients*, using standard gamble methodology, may be [more] relevant to this review' (AR, p. 182)	NICE guidance asks for a representative sample of the public as the source of utility data[36]. Further, there are specific concerns about the reliability of self-reports of children and adolescents with ADHD[59]
Quantitative synthesis (meta-analysis using mixed-treatment comparison technique)	For the base case economic model (primary evaluation), CGI-I based 'response rates' were synthesized and combined with utility estimates for responders and non-responders	Primary analysis resulted in *inconsistent rankings* of strategies (Tab. 6.7 of AR, p. 237), which were not even mentioned in the body of the text, except for the remark that 'the difference in QALY gains between the alternative treatments strategies was very small' (AR, p. 236)
	For secondary extensions of the model, response rates, which had been derived from heterogeneous criteria, were synthesized. Scales included (besides CGI-I and CGI-S) the ADHD-RS and the SNAP-IV (AR, p. 254), which were described as 'disease specific instruments' measuring 'health-related quality of life in children' (AR, p. 176)	The ADHD-RS and the SNAP-IV are typical narrow-band symptom scales[55,56] The assessment protocol[43] had stated that 'relative risks will only be pooled when this is statistically and clinically meaningful.'
	Heterogeneity of parameters such as patient populations (age, sex, comorbidity, etc.), study designs (efficacy, effectiveness), treatments (intensity, combination with non-drug treatment), and effect measures was mentioned repeatedly	Despite heterogeneity of effect measures across treatments and studies, there is no evidence that potential confounding effects between treatment strategies and effect measures was assessed

Table A3 *(continued)*

Subject	Statements and rationales	Consistency issue
Effectiveness versus efficacy distinction	The distinction between efficacy and effectiveness was not addressed in the assessment report; both terms were apparently used interchangeably, without discrimination	In his authoritative textbook, the senior author of the assessment report explained the fundamental importance of this differentiation for a meaningful economic analysis[83]. He stated that 'clinical trials are artificial environments, and do not provide all the economic information needed by decision-makers'[83], and 'for economic evaluations to be relevant, they need to reflect the real-world conditions faced by the decision-maker'[176]
	Discussion of the MTA study in the AR ran over five pages in the clinical effectiveness review (AR, pp. 164ff.)	However, the extensive measures to ensure fidelity and adherence were not described[63,64], and the mediator analyses clearly showing the impact of treatment adherence on treatment response were missed[64]; cf. AR, pp. 167ff
	'The effect of compliance on response rates to MPH-IR and MPH-MR is reflected in the model' (AR, p. 250)	This approach ignored the difference between efficacy and effectiveness trials and the AR did not discuss the extensive literature on noncompliance in general and in ADHD
	Noncompliance was assumed to be a subset of nonresponse in RCTs (AR, p. 232: implying that 'double-blind, double-dummy trials' … 'capture the effects of compliance')	*In contrast*, the senior author of the assessment stated elsewhere, 'great efforts are typically made in the conduct of a clinical trial to ensure that patients consume their prescribed medications' and, referring to the situation outside trials, '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]. He explicitly recommended the review of Hughes *et al.*[86] who concluded 'that *sensitivity analysis* should be applied appropriately to ascertain the impact of noncompliance on the cost-effectiveness of drug therapies.'
	'The exploration of the effects of non-compliance would involve a number of assumptions … it was felt that these modelling assumptions would not be reasonable given the lack of appropriate data, which would render the results of any *sensitivity analysis* around compliance uninformative to decision-makers' (AR, p. 233)	
	Any assumption was felt unjustified concerning 'the distribution of reduced compliance between morning, lunchtime, and evening doses of medication' (AR, p. 232f.)	A study from Canada had been in the public domain that indicated that MPH doses were frequently missed, with the second and in particular the third daily doses most affected[108]. This study was not discussed in the AR
Economic model	Stochastic analysis: 'The model is probabilistic, meaning that relevant input parameters are entered as probabilistic distributions rather than point estimates in order to *represent the uncertainty* around each point estimate' (AR, p. 220). 'The output from the model incorporates the uncertainty around the estimated response rates…' (AR, p. 229). As to the MTA subgroup analyses, it was stated that these evaluations 'should be seen as 'exploratory', because of the danger of *repeated statistical testing*	This is inconsistent with the use of CGI-I scores as primary efficacy parameter, because these were not the primary outcome parameters of the underlying RCTs. While it is quite legitimate to carry out secondary analyses, these should not be presented as main results *implied to fully capture uncertainty*
	'The nature of the treatment received in the community comparison arm of the MTA trial is still unclear, and as a result this data is omitted from the analysis' (AR, p. 254)	A table in the assessment report however states that three of its four arms were included: 'results for behavioral treatment were omitted as not relevant to this review' (AR, Tab. 6.17, p. 254). Thus it remains enigmatic which arm of the MTA was actually omitted from analysis
	For secondary analyses, 'we can also incorporate the results of the MTA trial, but only by assuming that the medical management group in that trial represents treatment with MPH-IR' (AR, p. 253)	It was noted in the effectiveness review (AR, p. 165) that 'most of the children in the community care group (97/146) received stimulant medication.'
		Using the medication management arm of the MTA study as a proxy for MPH-IR contradicts the importance of treatment adherence found in that study (see above[64]) as well as the extensive measures to ensure treatment fidelity and adherence in this trial[63,64]

Table A3 *(continued)*

Subject	Statements and rationales	Consistency issue
	'A number of studies excluded from the effectiveness review, for reasons of data presentations, were nevertheless … included in the calculation of response rate for the cost-effectiveness analysis. Further details of these excluded studies are given in appendix 3' (AR, pp. 225ff.)	Appendix 3 of the AR lists all studies excluded and does *not* provide any information, which studies might have been included in the economic model in addition to five studies of the clinical effectiveness review and Sharp *et al.*[75] (cf. AR, pp. 333ff.)
Extrapolation over 12 years	'In this extended analysis, costs are discounted at an annual rate of 6%, and health benefits are discounted at an annual rate of 1.5%, in accordance with NICE guidance' (AR, p. 233)	NICE guidance[36] specifies that costs and health benefits should be discounted at an annual rate of 3.5%
	'There is little data on long-term efficacy […] associated with medical management of ADHD' (AR, p. 19)	Inclusion of 14-months data from the MTA study remained enigmatic (see above, *Economic Model*), and other long-term studies were excluded from the review[57,153,154]
	Long-term sequelae of the disorder briefly mentioned as potential 'long-term benefits of treatment' (AR, p. 247)	Long-term sequelae of ADHD were neither mentioned in Executive Summary nor in Conclusions of the assessment report, and were not explored by literature search. Long-term extrapolation models did include a discussion of mid- to long-term effects ('sensitivity to time horizon': AR, pp. 245ff.)
Limitations	'Clear' conclusions, stated repeatedly (cf. below, Executive Summary): AR, pp.19, 261, 266	Caveats scattered throughout the report (e.g., AR, pp. 45, 224, 261ff.). Porzsolt *et al.*[177] identified 'a serious problem' in 'HTA reports which … express limitations in the discussions (read by many scientists) but not in the conclusions (read mainly by policy-makers).' See also point below
Executive Summary on Cost-effectiveness (AR, pp. 18f.)	'For a decision taken now, with current available data, the results of the economic evaluation clearly identified an optimal treatment strategy. That is, …'	Limitations of the model described incompletely, suggesting limited available information without indicating impact of study selection criteria chosen (e.g., regarding long-term studies excluded from review and economic model – see above)
	Caveats: 'The model is not without limitations. As identified in the clinical effectiveness review, the reporting of studies was poor, there is little data to discriminate between the drugs in efficacy or adverse events and there is little data on long-term efficacy and adverse events associated with medical management of ADHD. The data do not allow discrimination between patients with ADHD in terms of ADHD subtype, age, gender or previous treatment.'	Model however was limited to response rates based on CGI-I ratings, which were subsequently pooled for secondary 'sensitivity' analysis with various (heterogeneous) 'response rates'

Independent of their respective interpretation, each of the inconsistencies identified constitutes a gap of the assessment, i.e. important aspects were not adequately considered. AR: assessment report[27]

CrossRef links are available in the online published version of this paper:
http://www.cmrojournal.com
Paper CMRO-3661_ESD, *Accepted for publication*: 31 August 2007
Published Online: 08 January 2008
doi:10.1185/030079908X260817