1 **Title:**

2 A thematic analysis of the strengths and weaknesses of manufacturers' submissions to the NICE

3 Single Technology Assessment (STA) process

4

5 **Authors:**

6 *Christopher Carroll[a], Eva Kaltenthaler[a], Patrick FitzGerald[a], Angela Boland[b], Rumona Dickson[b]

7

8 **Address and affiliations:**

9 [a] Health Economics and Decision Science (HEDS), School of Health and Related Research

10 (ScHARR), University of Sheffield, Sheffield, S1 4DA, UK

11 [b] Liverpool Reviews and Information Group (LRiG), University of Liverpool, UK

12

13 * Corresponding author:

14 Christopher Carroll, Research Fellow, HEDS, ScHARR, University of Sheffield, Regent Court,

15 Regent Street, Sheffield, S1 4DA, UK

16 Email: c.carroll@sheffield.ac.uk

17 Tel: (0)114 2220864

18 Fax: (0)114 2724095

19

20 Other authors:

21 e.kaltenthaler@sheffield.ac.uk

22 p.fitzgerald@sheffield.ac.uk

23 a.boland@liverpool.ac.uk

24 r.dickson@liverpool.ac.uk

25

26

27

28

29

30

31

32

Abstract:

Objectives: The NICE single technology appraisal (STA) process in the UK has been underway for five years. Evidence Review groups (ERGs) critically appraise submissions from manufacturers on the clinical and cost effectiveness of new technologies. This study analysed the ERGs' assessment of the strengths and weaknesses of 30 manufacturers' submissions to the STA process.

Methods: Thematic analysis was performed on the textual descriptions of the strengths and weakness of manufacturer submissions, as outlined by the ERGs in their reports.

Findings: Various themes emerged from the data. These themes related to the processes applied in the submissions; the content of the submission (eg. the amount and quality of evidence); the reporting of the submissions' review and analysis processes; the reliability and validity of the submissions' findings; and how far the submission had satisfied the STA process objectives.

Conclusions: STA submissions could be improved if attention were paid to transparency in the reporting, conduct and justification of review and modelling processes and analyses, as well as greater robustness in the choice of data and closer adherence to the scope or decision problem. Where this adherence is not possible, more detailed justification of the choice of evidence or data is required.

MeSH keywords:

Review, systematic; Cost effectiveness; Cost Benefit Analysis; Pharmaceutical Industry; Drug approval

1   **INTRODUCTION**

2   The National Institute of Health and Clinical Excellence (NICE) is an independent organisation

3   which is part of the National Health Service (NHS) and responsible for providing guidance on the

4   promotion of good health and the prevention and treatment of ill health to England and Wales. One

5   of the key components of NICE's work involves technology appraisals which lead to

6   recommendations on the use of new and existing medicines and treatments within the NHS. NICE

7   technology appraisal guidance is mandatory in the NHS in England and Wales giving it the potential

8   to decrease variation in the provision of care across the nations.

9

10   The initial processes used to establish NICE guidance on technology appraisals are based on the

11   internationally-accepted models of reviewing clinical and cost effectiveness evidence. These include

12   a rigorous and systematic approach to identifying, evaluating and synthesising the available

13   evidence (clinical and cost data) carried out by groups of academic researchers (assessment groups)

14   aided by submissions from the involved manufacturers of the technologies. The result of this

15   synthesis is then considered by a carefully selected group of clinicians, health economists,

16   statisticians, patients and representatives from the NHS and the manufacturers, who together make

17   up the NICE Appraisal Committee (AC). This is known as the Multiple Technology Appraisal

18   (MTA) process.

19

20   A more rapid process became a political imperative and the newer Single Technology Appraisal

21   (STA) process was introduced for England and Wales in 2005.[1] The STA process was specifically

22   designed to appraise a new technology for a single indication, although there may be more than one

23   comparator and the process usually covers new technologies. Most importantly it was designed to

24   provide a more rapid appraisal process than the MTA process so that guidance for new products

25   could be produced as close to their launch into the NHS as possible. However, the MTA process

26   does not always take substantially longer to complete than the STA process.[1, 2, 3] The STA

27   process differs from the MTA process in that the manufacturer submission (MS) to NICE forms the

28   principal source of evidence for decision making. This is similar to the process followed in

29   Scotland.[4] The STA process is divided into stages.[5] Initially, provisional topics are identified

30   through a variety of sources and assessed by the NICE Topic Selection Committee. After formal

31   referral, NICE sets the STA timelines. Manufacturers are invited to prepare their submission to

32   NICE using a standard report template nine weeks after the scope is finalised. The MS is expected to

33   include a systematic review of the clinical effectiveness evidence for the technology under

34   consideration as well as a cost effectiveness analysis. Extensive guidance for manufacturers is

35   provided in the NICE guide to the methods for single technology appraisal.[6, 7] External

36   independent evidence review groups (ERGs), based in academic centres, are then charged with the

37   task of rapidly critically appraising the MS and identifying strengths, weaknesses and gaps in the

evidence presented. This work once again has to be completed within a set timetable (usually 8 weeks) and includes a clarification process, co-ordinated by NICE, in which ERGs can seek answers to questions of content and method that arise during their assessment of the submission. The resultant ERG reports form a part of the evidence considered by the AC when making a decision on the inclusion of the technology in current guidance.

The research reported here formed part of a broader project assessing the STA process, which involved detailing the process and outcomes for those STAs with documentation for that period (September 2006 - October 2009) and an analysis of the associated ERG reports and clarification letters.[2] The STA process was new and evolving at the time and this work was commissioned by the NIHR Evaluation, Trials and Studies Co-ordinating Centre (NETSCC) in order to detail the process, identify areas for improvement, and inform the development of a new template and guide for the production of manufacturer submissions and their assessment by ERGs. The aim of the particular work detailed in this paper was to analyse the strengths and weaknesses of the manufacturer submissions (MSs), as detailed by the report of the Evidence Review Groups (ERGs) assessing those submissions.

MATERIALS AND METHODS

A documentary analysis of the first 30 ERG reports produced for the STA process was undertaken. These reports were chosen because they were the only reports for which all documentation was available when the project started (March 2009). Attention was focused not on the context within which the documents were produced, nor on their subsequent impact on external decision-making processes, but rather exclusively on the content of the reports. The principal focus was the ERGs reported conclusions on the strengths or weaknesses of submissions. The criteria by which ERGs critiqued their respective submissions were not explicitly-defined *a priori*, but rather appeared to consist of applying the principles of standard methodological checklists such as QUOROM/PRISMA for the effectiveness review[8], Drummond and Jefferson (1996) [9] for the economic evaluation, and the basic outline of headings and brief definitions provided by the existing NICE ERG template. The emphasis may have altered or been adapted according to the scope of the particular technology being assessed. The 30 ERG reports were anonymised and none is referred to explicitly in this report.

The extraction of relevant data from the ERG reports was conducted by three team members (AB, CC, PF) using forms developed for this project and piloted on two ERG reports by all three authors. The aim of the extraction was to retrieve data on an ERG's critique of the strengths and weaknesses of a MS. For example, data were extracted on the ERGs comments on whether or not a meta-analysis

or sensitivity analysis (SA) should have been performed, or whether the ERG reported any errors in the model. Much of these data consisted of text, i.e. statements or summaries by the ERG. Thematic analysis[10] was the chosen method to analyse these data as it is grounded in the data and therefore permits the generation of a novel thematic framework reflecting the ERGs' assessments of the strengths and weaknesses of manufacturer submissions to the STA process. This method is interpretive and reductive; the first stage is data reduction, i.e. to reduce statements, comments, quotations or findings to a single theme, which captures or reflects those data. This interpretive process was initially performed by one reviewer (CC) on the extractions from a random sample of ten ERG reports. If any of the primary themes identified in this way was considered to be related then they were reduced further to a broader, meta-theme that captured them all. Definitions were then developed for each primary theme in order to produce greater reliability in the coding of data. Two members of the project team (RD, EK) then independently assessed whether these thematic interpretations of the data were both credible and appropriate, and whether the themes identified reflected the data. This led to a small number of revisions: the re-labelling of one primary theme; the reassignment of some data to different themes; and some further clarification of the themes' definitions. The extracted, textual data from the remaining 20 ERG reports were then coded using these agreed themes following a process akin to that described for framework analysis.[11] This was performed by two reviewers (CC, EK) for the remaining 20 ERG reports.

RESULTS

The thematic analysis generated a large number of primary and meta-themes (see **Figure**). Five meta-themes emerged, under which related primary themes could be meaningfully grouped. These five meta-themes related to the processes being applied in the manufacturer submissions; the reporting of the submissions' review and analysis processes (sometimes strong, sometimes poor); the submissions' satisfaction of objectives; the reliability and validity of the submissions' findings; and the content of the submission (e.g. the amount and quality of evidence contained in the submission).

Figure: Thematic framework based on analysis of data extracted

META-THEME	PRIMARY THEMES	DEFINITION
PROCESS	CONDUCT OF REVIEW CONDUCT OF MODELLING	How various methodologies have been applied, e.g. in the performance of the review, e.g. searching, screening, extraction, appraisal; in the modelling, in the meta-analysis, in the performance of the review.
	NO OR INADEQUATE ANALYSIS	Was an analysis performed? May include: failure to include or perform all necessary analyses (e.g. in a model), inadequate conduct of the review or analysis
	INAPPROPRIATE ANALYSIS	Was the analysis method applied appropriate? May include: inappropriate combining of data; inappropriate data being used to populate the model;
	ISSUES WITH DATA USED IN ANALYSIS	Issues concerning data used, especially in the modelling, including costs, parameters and assumptions
REPORTING	ADEQUATE REPORTING INADEQUATE REPORTING	Provision of sufficient or insufficient details about searching, selection, extraction, criteria, analyses performed and their rationale; descriptions and definitions provided in the Background section
SATISFYING OBJECTIVES	POPULATION ISSUES INTERVENTION ISSUES COMPARATOR ISSUES OUTCOME ISSUES NICE BASE CASE	Success or failure to answer question(s) set or for the submission to reflect the decision problem and its scope in terms of the target population, the intervention and its dose, relevant comparators and outcomes, and the NICE base case for the model
RELIABILITY & VALIDITY OF FINDINGS	UNCERTAINTY DUE TO ABSENCE OF EVIDENCE, POSSIBLE BIAS OR EXAGERRATED EFFECT	Excessive uncertainty surrounding results or model due to lack of evidence, bias within or across included trials or potential exaggerated effect of intervention; explicit concerns regarding validity
	RELIABILITY OF FINDINGS	Findings reported as being reliable (as opposed to uncertain as above) and not uncertain
	EXTERNAL VALIDITY	Issues affecting external validity, eg. specified differences between the trials and data and what exists in the UK, including population, dose, comparator, licensing, real world/current practice; future developments which may change key parameters
CONTENT	AMOUNT OF EVIDENCE	Weaknesses or strengths inherent in the trial evidence: Amount: concerns issues with number of trials included (eg. often only 1)
	QUALITY OF EVIDENCE	Quality: concerns how good the included evidence is, eg. very good, or very poor?

2

3

4

1   Table: Results

Meta-theme	Primary themes	Results	Illustrative quotations
Process	The clinical effectiveness review process	▪ Concerns with the quality of the searching, screening or quality assessment (11 reports); and the definition or application of inclusion criteria, especially for indirect comparisons (8 reports); ▪ The complete absence of a formal systematic review at all (2 reports)	"No clinical effectiveness or cost effectiveness reviews of the literature were included in the MS"
	The cost-effectiveness review and modelling process	▪ Failure to incorporate or capture adequate levels of uncertainty (3 reports); ▪ Calculation errors in the model requiring revision or correction (3 reports); ▪ Other technical, structural or design errors in the model or analyses (11 reports); ▪ Failure to control for confounders or queries over a key assumption, missing data or the absence of a Sensitivity Analysis (SA) or Probabilistic Sensitivity Analysis (PSA) (6 reports)	"Revisions to the model by the ERG render [technologies] as not cost-effective" "The approach taken to model the disease is pragmatic given the available data and previous MS models" "The PSA did not capture all of the uncertainty present in the decision"
	No analysis or inadequate analysis	▪ Failure to address heterogeneity of trials (1 report); ▪ The presence of errors in or the failure to perform relevant meta-analyses (3 reports); ▪ Failure to control for confounders or perform a key sub-group analysis (3 reports); ▪ Failure to perform a relevant analysis, e.g. a meta-analysis or SA or probabilistic sensitivity analysis (PSA) (8 reports); ▪ No validation of the model (2 reports); ▪ Inadequate analysis of safety outcomes (1 report)	"There is some heterogeneity between trials and this is not addressed" "subgroup analyses had not been adequately considered" "Model does not reflect real world decisions; parameter uncertainty is not given sufficient consideration" "The [meta]analysis contained a calculation error"
	Inappropriate analysis	▪ The combination, pooling or comparison of effectiveness data was viewed as being highly questionable (9 reports) ▪ Methods employed for the reported meta-analysis (8 reports) ▪ Methods employed for the reported modelling (4 reports)	"The methods used to pool data are inappropriate" "The conclusions of the indirect comparison are based on a visual not a statistical comparison of efficacy outcomes" "The validity of including unpublished *post hoc* analysis for two subgroups is questionable as both are likely to be underpowered"
	Issues with data being analysed	▪ The efficacy data being used in both direct and indirect comparisons (8 reports); ▪ The cost data used (7 reports); ▪ The utility or Quality of Life data in (5 reports); ▪ The data on population, comparators, outcomes or various model parameters (10 reports); ▪ The use of unpublished data (3 reports); ▪ The data being used in the s SA or PSA (2 reports)	"For some model parameters the authors had employed standard deviations (measures of sample dispersion) rather than standard errors" "The second pivotal trial does not disaggregate outcome data for the relevant population, data for all populations are analysed" "The majority of the reference data presented in the MS were not fully published"

Meta-theme	Primary themes	Results	Illustrative quotations
Reporting	Inadequate reporting	▪ Poor descriptions of the searches undertaken, prohibiting replication, for both clinical and cost-effectiveness reviews (8 reports); ▪ Lack of transparency of the clinical review processes generally, both for direct and indirect comparisons (e.g. screening, extraction and quality appraisal) (7 reports); ▪ Concerning the number of included studies and their characteristics for both clinical and cost-effectiveness reviews (8 reports); ▪ Failure to report data and data sources in full (3 reports); ▪ Failure to describe all of the methods being employed in direct or indirect comparisons (8 reports); ▪ Failure to provide adequate descriptions of the analyses more generally (3 reports); ▪ Poor reporting of included studies and data in tables (3 reports); ▪ Background section deemed inadequate and lacking key information (5 reports); ▪ Failure to describe adequately either the parameters or assumptions behind the model, the generation or source of various values, or the impact of bias from (11 reports)	"There was a lack of annotation or documentation associated with the original model" "The methods used by the manufacturer lack transparency" "The results of data pooling and the [mixed treatment comparison] MTC should be treated with caution as the methods are not fully described" "The ERG would have preferred more information on average length of disease-free survival and current service provision" "The processing of data in models is not clearly documented" "A number of assumptions were not fully justified and possibly introduced bias into the results" "The reporting of the included RCT lacks detail and clarity making interpretation of clinical effectiveness results difficult"
	Adequate reporting	▪ The searches for the effectiveness review was praised (4 reports); ▪ The description of the model and its data sources praised (5 reports); ▪ Good reporting of statistics, and a well-reported multiple treatment comparison (MTC) process (1 report)	"generally good statistical reporting" "The MS provides a generally accurate and thorough discussion of the background to the disease … and its treatments"
Content	Amount	▪ Submission was based on the evaluation of one or two trials only (10 reports); ▪ The absence of any head-to-head trials of relevant technologies (3 reports)	"There does not appear to be sufficient evidence to compare the relative effectiveness of [technology] and [technology]" "There is no direct evidence to support the use of the intervention in the key subgroup of patients"
	Quality	▪ ERGs explicitly report trials as being of good or reasonable quality (9 reports)	"The four included trials were of reasonable methodological quality" "The main RCT in the MS was well-conduced when assessed using the NICE internal validity criteria"

Meta-theme	Primary themes	Results	Illustrative quotations
Satisfying Objectives	Population	■ Differences between the population defined in the decision problem and the population being evaluated in the submissions trials (11 reports); ■ Differences between the UK population and the trial population based on age, treatment or current practice (6 reports); ■ Differences between the licensed population and the trial population (2 reports); ■ Failure to consider the effect of the treatment on different, relevant sub-groups of patients (6 reports); ■ Problems with the definition of the population in the submission (4 reports)	"It is uncertain whether the trial population is sufficiently similar to the UK" "Some potentially relevant populations were excluded from main trial" "The trial population does not reflect the licensed population" "There is a subtle but important variation in the description of the target population between the scope and the MS"
	Intervention	■ The definition of the intervention was an issue, i.e. the inclusion or failure to include the intervention as combination therapy or monotherapy (5 reports); ■ Differences between the trial interventions and UK practice (2 reports); ■ Differences between the licensed intervention and the trial intervention (1 report); ■ Dose being evaluated was reported as an issue (1 report)	"Scope covers both mono-therapy and treatment in combination - the latter is not included in the decision problem" "Definition of the intervention in the MS is in accordance with the proposed marketing authorisation" "There is uncertainty regarding the dose used in the key trial"
	Comparators	■ Failure to consider one or more of the comparators designated in the decision problem, or the submissions' use of a combination of comparators not admitted in the decision problem (13 reports); ■ Comparator being evaluated was not in use in the UK or that a non-optimal dose of the comparator was being assessed (3 reports); ■ Lack of definition for the chosen comparator (2 reports); ■ Non-optimal treatment duration for the comparator in the model (1 report); ■ Mixed-treatment comparison (MTC) conducted using comparators other than the principal comparator named in the scope (1 report)	"This evidence cannot be used directly to answer the questions raised in the decision problem because in the main trial [technology] was not compared to a relevant comparator" "The MS did not clearly state which technologies were the key comparators" "The choice of comparator reflects the comparator in the [trial] rather than possible relevant comparators in an NHS context"
	Outcomes	■ Uncertainties regarding the appropriateness of the outcome being measured (6 reports); ■ Lack of clarity on how Quality of Life was being measured (2 reports); ■ Inadequacy of the safety measures presented (1 report); ■ Concerns about a trial's lack of power to detect the secondary outcomes being presented (1 report)	"appropriate and clinically meaningful outcomes" "MS does not specify how three outcomes are to be measured" "Inadequate consideration of safety (greater toxicity and long term side effects) of [technology]"
	NICE base case	■ Multiple deviations from the base case, or reported specific deviations in relation to either the calculation of utilities or the comparator (8 reports)	"The [cost effectiveness analysis] is in accordance with the NICE reference case and the scope"

Meta-theme	Primary themes	Results	Illustrative quotations
Reliability and validity	Uncertainty due to absence of evidence, possible bias or exaggerated effect	▪ Possible exaggerated effect of the technology in the analysis especially the relative efficacy of the technology versus relevant comparators (12 reports); ▪ Uncertainty concerning the safety of the technology (1 report); ▪ Uncertain levels of risk for different populations (1 report).; ▪ Uncertainties due to issues with the parameters or values in the model (10 reports); ▪ Exaggeration or over-estimate of benefits or costs (7 reports); ▪ Excessive extrapolation from the data (4 reports); ▪ Uncertainty due to the model's high degree of sensitivity to values or assumptions within the model (6 reports) ▪ Uncertainties regarding relative efficacy of technologies as no head-to-head trial had been performed or because the findings were based on only a single trial (4 reports)	"Benefits of the technology on rates of recurrence are unknown beyond 3 to 4 years" "In the single trial, the primary outcome only became significant in favour of the intervention after a post-hoc adjustment had been made to correct for a 'clinically relevant' imbalance between trial arms" "A number of assumptions were not fully justified and possibly introduced bias into the results" "The efficacy of one of the drugs is still under investigation" "There are issues concerning indirect comparison which may over-estimate the effect of [technology]"
	Reliability	▪ The submission was deemed to offer a convincing case for the technology, an unbiased estimate of treatment effect or a fair interpretation of the trial data (9 reports); ▪ The submission was deemed to present a reasonable estimation of the cost-effectiveness of the technology versus relevant comparators, that the modelling of the costs of the technology appeared sound, or that the model was superior to previous published models (3 reports)	"The ERG does not consider the manufacturer's PSA to be reliable as it excludes some important variables and fails to account for potentially important parameter covariances" "The MS represents a reasonable estimation of the cost-effectiveness of the drug against relevant comparators" "The results represent an unbiased estimate of the treatment effect"
	External validity	▪ Differences between the trial and UK populations (6 reports), including sub-groups of UK patients likely to receive the treatment (1 report); ▪ Differences between the treatment practices being evaluated in trials and clinical practice in the UK or Europe (7 reports); ▪ Differences between the trials and "real world" clinical practice generally (2 reports) ▪ Criticised for use of non-UK sources of data for the model (2 reports)	"Trial populations may not be generalisable to UK" "Trial results may not be directly applicable to the target population" "Disease free and overall survival may differ in the UK from the comparator arm in the model" "The treatment does not reflect stated NICE scenarios or clinical practice in Europe" "The model does not reflect real world decisions"

1
2

1 Process

2 The meta-theme of process reflected the primary themes of how well or how badly the

3 manufacturer conducted the review and the modelling processes; the inadequacy of the

4 analyses or the appropriateness of the analyses performed and reported in the submission;

5 and the ERGs' reported concerns with the data that were used in the analyses. On many

6 occasions, elements of the systematic review process, such as searching and data extraction,

7 or the structure of the model, prompted positive comments from an ERG. However, more

8 than half of the ERG reports explicitly criticised the conduct of the clinical effectiveness

9 review and/or the model (see **Table**). The analyses performed by the manufacturers and

10 reported in the submission constitute another element of the effectiveness review process that

11 was frequently critiqued by the ERGs, but it was not simply the failure to perform a

12 necessary analysis, but rather the performance of an inappropriate analysis also that

13 generated criticism in a majority of ERG reports. Finally, two thirds of ERG reports (20/30)

14 contained criticisms relating to the data being used in analyses and models.

15

16 Reporting

17 This meta-theme was derived from explicit comments made in ERG reports on the quality of

18 the manufacturers' description of the conduct of both the reviewing and the modelling. Only

19 in a small number of cases did ERGs make explicit, positive comments on the description of

20 the technology or processes within manufacturers' submissions. Nine ERGs commented on

21 the accuracy and adequacy of the background section provided in the submission, albeit in

22 four cases stating that further information would still have been useful. It was far more

23 common however for ERGs to perceive as inadequate the reporting or description of the

24 processes being undertaken or the sources of data being used. A total of 27 out of 30 ERGs

25 reported some form of inadequacy in the reporting of a submission.

26

27 Satisfaction of objectives

28 This theme emerged from issues surrounding the relationship between the decision problem

29 and the scope issued by NICE, and whether or how far the elements of these were

30 satisfactorily addressed in the manufacturer submissions. The five primary themes that

31 generated this meta-theme relate to the objectives determined by the decision problem, i.e.

32 the population, intervention, comparator, outcomes and the NICE base case.

33

34 Twenty of the thirty ERG reports did raise issues with the trial populations being presented

35 and considered in the submission. The second major disparity between submissions and the

36 requirements of the decision problem concerned the comparators. Twelve ERG reports found

37 nothing to criticise in this regard, but eighteen did raise various issues with the comparators

considered in submissions. The principal issue, raised in thirteen reports, involved a submission's failure to consider one or more of the comparators designated in the decision problem, or the submissions' use of a combination of comparators not admitted in the decision problem. By contrast, few ERGs commented on a submission's failure to satisfy the intervention or outcome elements of the decision problem. Twenty-one ERG reports reported no problem with the intervention being evaluated in the submission. Twenty-one reports also did not have any criticism to make of the outcomes presented in the submissions. The principal issue in six of the remaining reports concerned uncertainties regarding the appropriateness of the outcome being measured. The vast majority of the issues relating to the satisfaction of objectives were raised in the appraisal of the clinical effectiveness evidence. Very few, for example, only those relating to the measurement of Quality of Life as an outcome, were issues raised principally or exclusively in the cost-effectiveness sections of the reports. In relation to satisfying objectives, the principal focus of the cost-effectiveness section concerned the submissions' failure or otherwise to satisfy the requirements of the NICE base case when developing the economic model. The majority of submissions appear to have adhered to the NICE base case scenario and prompted no criticism: only eight ERGs reported any issues.

Reliability and validity

The reliability and validity theme emerged from four primary themes that were interpretations of ERG comments on the robustness or limitations of the submissions' findings. These themes were: uncertainty due to the possibility of bias or an exaggerated estimate of effect; uncertainty due to the absence of evidence; the reliability (rather than uncertainty) of the findings; and external validity, i.e. how far the ERG considered the submission to be externally valid for the intended population and service. The level of uncertainty surrounding the findings presented by the submission was a frequent cause of criticism in ERG reports. Twenty-seven out of thirty ERG reports stressed the presence of bias within the analyses, thus highlighting the lack of certainty surrounding the results presented in the submissions. Uncertainties within the clinical effectiveness analyses impacted on the models, which were also subject to other biases, such as uncertainties due to issues with the parameters or values in the model. By contrast, few ERG reports stated that it was the lack of evidence rather than the quality of the evidence or its analysis that generated uncertainty in the submissions' findings. ERG reports did sometimes also explicitly comment that the findings of the effectiveness or cost-effectiveness analyses were strong and reliable, but this was relatively infrequent. Finally, seventeen ERG teams explicitly queried the external validity of the submissions' findings.

1 <u>Content</u>

2 The ERGs often commented on the amount and quality of the trial evidence that formed the

3 basis for the submission as this affected both the internal and external validity of the

4 submission and its reviews. However, unlike the other themes, this issue is not addressed

5 readily by the improved conduct of reviews, analyses, modelling or reporting of the

6 processes used. Nevertheless, it is an element of the submission that attracted comment from

7 ERGs and so is represented in this analysis. The limited amount of relevant trial evidence,

8 for example, the fact that the submission may be based on the evaluation of a single RCT, or

9 perhaps two such trials only, was sometimes explicitly raised as a point by ERGs, as was the

10 quality of the included trials was also often commented on by ERGs, as this had implications

11 for the validity of the review, model and submission. However, the included trials, even if

12 there was only one, were more often explicitly reported as being of good or reasonable

13 quality. Small sample sizes in trials, and limited follow-up, were other factors affecting the

14 amount and quality of the evidence that drew comments from three ERGs.

15

16 **DISCUSSION**

17 The role of the ERG team is to *critically appraise* the MS, so emphasis is on identifying

18 aspects where there are concerns rather than highlighting occasions where something has

19 been undertaken appropriately. More than half of the ERG reports explicitly criticised the

20 conduct of the systematic review within the MS. Other criticisms covered both failure to

21 perform a necessary analysis and the performance of an inappropriate analysis. However half

22 of all ERG reports did pass positive comments on the appropriateness of the structure of the

23 economic model presented or the reasonableness of the modelling.

24

25 It may be argued that new technologies may often be unable to provide extensive data for

26 analysis. However it is possible to address issues raised by the absence for example of

27 sufficient direct comparison trial data by using available multiple treatment or indirect

28 comparison methods, such as network meta-analysis. The findings indicate that

29 manufacturers often chose inappropriate methods or did not provide sufficient detail for the

30 ERG to understand the approaches being used. Two thirds of ERG reports also contained

31 criticisms related to the data being used, especially the data employed in the cost-

32 effectiveness model. The NICE requirements for the economic model and analyses, such as

33 probabilistic sensitivity analyses, may be deemed "excessive", but manufacturers are aware

34 of these requirements and there is prior evidence that industry can be optimistic in their cost

35 analyses.[12, 13].

36

There were major issues regarding poor reporting of processes used in manufacturer submissions. These included poor descriptions of literature searching, and lack of transparency in the description of the processes used for both the clinical and cost effectiveness sections of the submissions, with reporting of the cost effectiveness model being the most common criticism by ERGs. The population and comparator represented the key items in the decision problem that ERGs' assessed as being either poorly or inadequately addressed by manufacturers. It is also an issue with implications for the ERGs' critique. A lack of concordance between the scope and the published and unpublished research does not make the decision problem inappropriate: the question posed is valid for policy-makers; and the manufacturers' trials will have been conducted in particular contexts and jurisdictions for particular reasons. A lack of "fit" between the question and the available research may therefore be unavoidable in some cases, but it does have implications for the external validity of the results and how far they can be said to apply to the context of interest, i.e. the UK NHS. For this reason, the failure to satisfy the objectives is important. It is important to note that this analysis is based on what was reported in the ERG reports after the clarification letter stage, in other words, after the manufacturers' have already had the opportunity to respond to any criticisms highlighted by the ERG. Indeed, manufacturers may have satisfied many requests for information at this interim stage. The ERG reports described here therefore represent the final critique.

Some of the themes also relate either directly or indirectly to one another. An ERG's assessment of the deficiencies in the processes performed in a submission, e.g. the failure to perform an appropriate analysis, the presence of bias within analyses, or issues with the data used, all also directly impact on its assessment of the reliability and validity of the findings of the submission, another key theme. In the same way, the failure of submissions to address or satisfy the objectives as outlined in the decision problem directly influenced ERGs' assessments of the external validity of the findings as they related to clinical practice in the UK. This suggests that if a submission addresses issues relating to the processes being used, and presents clear rationale for the choices being made, and any unavoidable limitations of evidence or analysis, then ERG assessments of the reliability and validity of the submissions' findings will be positively affected also.

The only other previously published analysis of similar submissions, the economic analyses submitted to the Australian Pharmaceutical Benefits Scheme, identified some similar problems: uncertainties surrounding the estimates of effect, and issues with the parameters, values and structure of the economic models, and errors within the models.[14] Choice of

comparator was also highlighted as an issue. This earlier, Australian evaluation reported that approximately two thirds of the problems were deemed to be avoidable. These findings have some strong similarities with the results of the present analysis. It has also been commented elsewhere that transparency of both the methods required and the criteria by which the submissions are to be judged should improve manufacturer submissions.[3] The findings of this paper suggest however that manufacturers do not always satisfy these criteria as they are outlined in the methods guide and templates.[6, 7] What may be required is even greater definition of these criteria, and greater stress on the need to prepare submissions with the methods guide and template in mind. Finally, eighteen of the thirty STAs discussed here were approved at the Final Appraisal Determination (FAD). A further three were approved after appeal; one was terminated and the remaining eight were not approved. This indicates how technologies may be approved even when there are issues with the submission either because those issues might not be deemed "vital", because they have been adequately addressed by the ERG, or because other variables also help to shape the decisions. However, the ERGs' assessments were obviously also instrumental.

The limitations of this study are that it focuses on ERG perceptions of the weaknesses of the manufacturers' submissions, as critique was the function of the ERGs and their reports. Although certain appraisal criteria were being followed, these were not standardised or absolute. Consequently, not every ERG therefore always reported the presence of a problem that was reported by another ERG. The issues deemed the most important by the ERGs doubtless attract the most comment. The appraisal process was and is still evolving. This analysis therefore offers a map of the key issues raised by ERGs' evaluation of submissions, rather than a highly specific analysis focusing on, for example, the modelling approaches adopted. There may be value in such future work. However, the internal validity of the findings reported here is enhanced by the consistency with which the identified themes emerged from across substantial numbers of the reports. Given that the number of STAs now total around one hundred [15], there may also be some value in conducting a similar assessment of a sample of more recently completed ERG reports to compare findings. The results may be the comparable however because no changes have been made yet to the methods guide or criteria.[6, 7] Alternatively, as individual manufacturers also garner more and more experience of both the process and the focus of ERG criticisms, then the perceived "quality" of some submissions may indeed have changed.

CONCLUSION

The ERGs appear to be performing an intensive evaluation of manufacturers' submission to the STA process using the NICE methods guide and template, and applying standard principles of critical appraisal for systematic review and cost-effectiveness models. The ERG criticisms may be reduced if manufacturers address the issues raised by this analysis of assessments of previous STA submissions, i.e. transparency in the reporting, conduct and justification of review and modelling processes and analyses, as well as greater robustness in the choice of data and closer adherence to the scope or decision problem. In cases where this is not possible, more detailed justification of the choice of evidence or data might also temper the criticism of submissions. The responsibility on both parties is substantial given the potential implications for health services provision that emerge from the process.

1 **REFERENCES**

2 [1] Drummond M, Sorenson C. Nasty or Nice? A perspective on the use of Health

3 Technology Assessment in the United Kingdom. *Value in Health* 2009; 12(2): S8-13.

4

5 [2] Kaltenthaler E, Boland A, Carroll C, Dickson R, FitzGerald P, Papaioannou D. Critical

6 Appraisal of Manufacturer submissions to the NICE STA process. *Health Technology*

7 *Assessment* 2011 (in press).

8

9 [3] Barbieri, M, Hawkins, N, Sculpher, M. Who does the numbers? The role of third-party

10 technology assessment to inform health systems' decision-making about the funding of health

11 technologies. *Value in Health* 2009; 12: 193-201.

12

13 [4] Cairns J. Providing guidance to the NHS: the Scottish Medicines Consortium and the

14 National Institute for Clinical Excellence compared. *Health Policy* 2006; 76: 134-43.

15

16 [5] National Institute for Health and Clinical Excellence. Guide to the single technology

17 appraisal process. October 2009. NICE: London;

18 http://www.nice.org.uk/media/913/06/Guide_to_the_STA-proof_6-26-10-09.pdf   (accessed

19 17 September 2010).

20

21 [6] National Institute of Health and Clinical Excellence. Single Technology Appraisal (STA).

22 Specification for the manufacturer/sponsor submission of evidence. October 2009. London.

23

24 [7] National Institute of Health and Clinical Excellence. Guide to the Methods of Technology

25 Appraisal. Updated. June 2008. London.

26

27 [8] Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting

28 Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *British Medical*

29 *Journal* 2009; 339: b2535.

30

31 [9] Drummond M, Jefferson T. Guidelines for authors and peer reviewers of economic

32 submissions to the BMJ. *British Medical Journal* 1996;313:275-83.

33

34 [10] Miles, M and Huberman, A. *Qualitative data analysis: An expanded sourcebook*. 2nd ed.

35 Sage; 1994.

36

1   [11] Ritchie, J and Spencer, L. Qualitative data analysis for applied policy research. In:

2   *Analyzing Qualitative Data*, ed. Bryman A, Burgess R. Routledge; London & New York;

3   1994: 173-194.

4

5   [12] Chauhan, D, Miners, A, Fischer, A. Exploration of the difference in results of economic

6   submissions to the National Institute for Clinical Excellence (NICE) by manufacturers and

7   assessment groups. *International Journal of Technology Assessment in Health Care* 2007; 23:

8   96-100.

9

10   [13] Miners, A, Garau, M, Fidan, D, Fischer, A. Comparing estimates of cost-effectiveness

11   submitted to the National Institute for Clinical Excellence (NICE) by different organisations:

12   Retrospective study. *British Medical Journal* 2005; 330: 65.

13

14   [14] Hill, S, Mitchell, A, Henry, D. Problems With the Interpretation of Pharmacoeconomic

15   Analyses. A Review of Submissions to the Australian Pharmaceutical Benefits Scheme.

16   *Journal of the American Medical Association* 2000; 283: 2116-2121.

17

18   [15] Kaltenthaler , Papaioannou d, Boland A, Dickson R. The NICE Single Technology

19   Appraisal Process: Lessons from the first four years. *Value in Health* 2011; (forthcoming).