
Article:
Carroll, C., Kaltenthaler, E. orcid.org/0000-0002-6185-8321, Hill-McManus, D. et al. (6 more authors) (2017) The type and impact of Evidence Review Group (ERG) exploratory analyses in the NICE Single Technology Appraisal (STA) process. Value in Health, 20 (6). pp. 785-791. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2016.08.729

Our reference: JVAL 1623 P-authorquery-vx

AUTHOR QUERY FORM

ELSEVIER	**Journal: JVAL** **Article Number: 1623**	**Please e-mail your responses and any corrections to:** **E-mail: a.rodenmeyer@elsevier.com**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the Q link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
Q1	Please confirm that given names and surnames have been identified correctly and are presented in the desired order.
Q2	Author: • If there are any drug dosages in your article, please verify them and indicate that you have done so by initialing this query. • Please review and confirm the accuracy and completeness of any affiliations.
Q3	AU/PE: Should "STA" in the article title be spelled as "Single Technology Appraisal" instead of "Single Technology Assessment" (i.e., is it "Appraisal" instead of "Assessment")? It has been used with the word "appraisal" throughout the main text.
Q4	AU: Please note that Table 2 outlines 6 (and not 7) parts of the data extraction tool. Please revise accordingly.
Q5	AU: The sentence "…72 STAs had received a positive recommendation, so that twenty nine STAs…" was unclear as given. Kindly check whether the edited sentence conveys the intended meaning.
Q6	AU: In the Table 3 footnote, please include the spelled out forms of MS, PFS, and PRIMA.
Q7	AU: In the Table 4 footnote, please include the spelled out forms of EOX and HCX.
Q8	AU: In the Table 5 footnote, please include the spelled out form of PFS.
Q9	AU: Please provide a heading for column 1 in Table 9.
Q10	AU: Please provide a heading for column 1 in Table 10.

Thank you for your assistance.

Please check this box or indicate your approval
if you have no corrections to make to the PDF file

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

The type and impact of Evidence Review Group (ERG) exploratory analyses in the NICE Single Technology Assessment (STA) process

Christopher Carroll, PhD[1,*], Eva Kaltenthaler, PhD[1], Daniel Hill-McManus, BSc[2], Alison Scope, PhD[1], Michael Holmes, MSc[1], Stephen Rice, MSc[3], Micah Rose, MSc[4], Paul Tappenden, PhD[1], Nerys Woolacott, PhD[5]

[1]University of Sheffield, Sheffield, UK; [2]Bangor University, Bangor, UK; [3]Newcastle University, Newcastle, UK; [4]University of Southampton, Southampton, UK; [5]University of York, York, UK

ABSTRACT

Background: As part of the UK National Institute for Health and Care Excellence (NICE) single technology appraisal process, independent evidence review groups (ERGs) critically appraise a company's submission relating to a specific technology and indication. **Objectives:** To explore the type of additional exploratory analyses conducted by ERGs and their impact on the recommendations made by NICE. **Methods:** The 100 most recently completed single technology appraisals with published guidance were selected for inclusion. A content analysis of relevant documents was undertaken to identify and extract relevant data, and narrative synthesis was used to rationalize and present these data. **Results:** The types of exploratory analysis conducted in relation to companies' models were fixing errors, addressing violations, addressing matters of judgment, and the provision of a new, ERG-preferred base case. Ninety-three of the 100 ERG reports contained at least one of these analyses. The most frequently reported type of analysis in these 93 ERG reports related to the category "Matters of judgment," which was reported in 83 reports (89%). At least one of the exploratory analyses conducted and reported by an ERG is mentioned in 97% of NICE appraisal consultation documents and 94% of NICE final appraisal determinations, and had a clear influence on recommendations in 72% of appraisal consultation documents and 47% of final appraisal determinations. **Conclusions:** These results suggest that the additional analyses undertaken by ERGs in the appraisal of company submissions are highly influential in the policymaking and decision-making process.

Keywords: evidence review groups (ERGs), health policy, National Institute for Health and Care Excellence (NICE), single technology appraisal (STA).

Introduction

The National Institute for Health and Care Excellence (NICE) single technology appraisal (STA) process is undertaken for a technology for a single indication [1]. This process involves the submission of an evidence dossier by the manufacturer of the technology. One of nine independent evidence review groups (ERGs) then undertakes a critical appraisal of this submission and produces a report. As part of this process, the ERG might undertake additional analyses, so-called exploratory analyses, to explore uncertainties around the company's model and their implications for decision making. The number and type of these exploratory analyses vary between appraisals. The ERG reports are a central component of the evidence considered by the NICE Technology Appraisal Committees (ACs) in their deliberations. The findings of the AC are used to produce an appraisal consultation document (ACD) and, after further considerations and a consultation period, a final appraisal determination (FAD), which result in NICE guidance.

The aim of this research was to develop an understanding of the number and type of exploratory analyses undertaken by the ERGs within the NICE STA process and to understand how these analyses have been used by NICE ACs in their decision-making process. For the purpose of this research, an exploratory analysis was defined as any additional analysis generating an incremental cost-effectiveness ratio (ICER) and ~~was~~ included in the ERG report section as "Exploratory and sensitivity analyses undertaken by the ERG." This is most commonly reported as section 6 of an ERG report, on the basis of the suggested ERG report template. This study aimed to address the following objectives:

1. to identify exploratory analyses conducted by the ERG, as defined earlier;
2. to identify ERG approaches to the exploratory analyses of economic evidence submitted by companies for NICE STAs and to categorize these approaches by type of analysis performed; and

* *Address correspondence to:* Christopher Carroll, Health Economics and Decision Science, School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, Regent Street, Sheffield S1 4DA, UK.
E-mail: c.carroll@shef.ac.uk.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2016.08.729

3. to make an assessment of the degree to which the exploratory analyses influenced committees' considerations and recommendations.

This article should be read in conjunction with the monograph by Kaltenthaler et al. [2], which reports on the frequency of exploratory analyses in ERG reports and the extent to which they might be predicted by factors such as company, ERG group, and disease area. The recently published work of Ghabri et al. [3] has considered some similar issues with reference to the cost-effectiveness analyses in manufacturer submissions to the French National Authority for Health. The findings of the present study will also be considered in this context.

Methods

The 100 most recently completed STAs (since 2009) for which final guidance has been published were selected for inclusion in the analysis. A list of the STAs included in this analysis is provided in the appendix to the article by Kaltenthaler et al. [2]. The research required extraction of relevant data from more than 400 separate documents, which were made available to the project team by NICE. These included unredacted versions of the ERG reports (used by the ACs), the first ACD issued (subsequent ACDs were not considered), and the first FAD (when more than one FAD had been produced). A data extraction tool was formulated to extract relevant data to address the project objectives. STA reports, ACDs, and FADs all have a basic, standard structure, which facilitated data extraction. The ERG reports have a specific "Exploratory" or "Additional analyses" section, usually section 6, from which the data on exploratory analyses were extracted. Nevertheless, ERG reports can vary greatly in their level of detail and their description of analyses. A final agreed data extraction tool was developed after a piloting process. The categories of exploratory analyses to be used in this study were based on discussions with the whole project team and an existing relevant published taxonomy of errors and other threats to the credibility of health economic models [4]. This approach is similar to that of framework analysis techniques for developing an a priori framework for coding qualitative data [5]. The categories were defined to facilitate consistency of coding and were amended following the piloting process. The category "Matters of judgment" was originally composed of three more specific categories: 1) uncertainty and evidence variation, 2) alternative data, and 3) ERG subjective judgment. It was, however, found that the descriptions of the analyses in the reports were often too vague to ensure that the information was being interpreted and coded consistently into one of these specific categories. For this reason, a single, broader category of "Matters of judgment" was created. The final four agreed categories of exploratory analyses are listed in Table 1. This simple scheme facilitated consistency of coding between data extractors.

The seven parts of the data extraction tool are outlined in Table 2. All data extractions were double-checked by at least two researchers. Some of the data extraction was simple and straightforward, but researchers sometimes had to exercise interpretation for data relating to exploratory analyses, for example, whether specific analyses influenced AC recommendations. To address issues of interpretation such as this, the key data used in the synthesis were then reduced to whether just "one or more" exploratory analyses were explicitly cited as having an influence on a recommendation. Such decisions were taken to make the most of the data.

A narrative synthesis was performed on the extracted data from the 100 ERG reports [6]. This involved summarizing the key data through text and tables and then using narrative to highlight any potentially important patterns or relationships in the data. This approach was taken because the large number of reports and documents prevented meaningful, in-depth analysis of the text using qualitative methods, but the number of reports was not large enough to permit meaningful statistical analysis of the data. In conducting the synthesis, the following assumptions were made:

1. Every exploratory analysis had to have a separately reported ICER. If an analysis combined the results of two or more exploratory analyses to calculate the third ICER, for example, to create the ERG's preferred base case, then this was considered to be a further separate analysis.
2. The ICER was for the technology against its principal comparator or in the principal scenario (its most likely use in clinical practice).
3. When the base case or preferred ICER reported by the company, ACD, or FAD was a range or multiple ICERs (e.g., for subgroups or scenarios), then the lowest ICER was used.
4. If no ICER was reported but a technology was considered in the ACD or FAD to be "a cost-effective use of NHS [National Health Service] resources," then it was deemed to be so at a threshold of £20,000 per quality-adjusted life-year (QALY) gained, on the basis of the perceived importance of the £20,000 per QALY threshold for NICE decision making [7].
5. If the ACD or FAD simply stated that the technology was "dominated," then it was assumed that it was not cost-effective.
6. The data on whether the analyses are mentioned by the ACD or FAD, or influenced their recommendation, relate to the exploratory analyses described in the ERG report or a specific addendum document (rather than any analyses conducted between the ACD and FAD).

Table 1 – Summary of categories of exploratory analysis present in ERG reports produced for the NICE STA process.

Category	Definition
Fixing errors	The ERG considered that something was unequivocally wrong in the company's submitted model.
Addressing violations	The ERG considered that the NICE reference case, scope, or best practice had not been adhered to for one or more parameters or values, including missing out relevant key comparators, and hence the model was not fit for purpose.
Matters of judgment	The ERG did not consider that the submitted model was wrong as such, but amended the model by conducting an analysis (often a sensitivity or threshold analysis) to test uncertainties within the evidence or model, or because reasonable alternative assumptions could be applied. These could be hypothetical or based on alternative data in the published literature or provided by a company.
ERG-preferred base case	The ERG conducted its own specific preferred base-case analysis. This might be the result of a series of exploratory analyses. This base case might still not be ideal from the ERG's perspective.

ERG, evidence review group; NICE, National Institute for Health and Care Excellence; STA, single technology appraisal.

Table 2 – Summary of data extraction tool.

Section	Details of fields
Basic characteristics	• STA title • ERG • Company • Disease area • Prespecified subgroups or end-of-life criteria applied? Extractors selected from a prespecified list of variables for most fields.
Company's base-case ICER(s)	• What is the ICER for the single technology against the principal comparator(s)? • What are the ICERs for the technology against the next best nondominated or baseline comparators? • What type of model was submitted?
Number and type of exploratory analyses conducted by ERG	• See Table 1. Extractors selected from a prespecified list of categories. • Data extracted included the source(s) of alternative data, when reported.
ACD	• What was the preferred ICER and its source? • What was the recommendation (prespecified list of variables)? • Did the ACD mention one or more of the exploratory analyses and was an ICER mentioned? • Did one or more exploratory analyses influence the AC's recommendation (i.e., is the analysis or its ICER cited specifically in relation to the recommendation)?
Additional work	• Was additional work undertaken by the company and/or ERG between the ACD and the FAD?
FAD	• What was the preferred ICER and its source? • What was the recommendation (prespecified list of variables)? • Did the ACD mention one or more of the exploratory analyses and was an ICER mentioned? • Did one or more exploratory analyses influence the AC's recommendation (i.e., is the analysis or its ICER cited specifically in relation to the recommendation)?

AC, Appraisal Committee; ACD, appraisal consultation document; ERG, evidence review group; FAD, final appraisal determination; ICER, incremental cost-effectiveness ratio; STA, single technology appraisal.

7. If any work was conducted by a company between an ACD and an FAD, it was assumed that it was critiqued by the ERG as a standard procedure (explicit requests by an AC for an ERG to conduct such work were rarely recorded in the ACD).
8. Evidence of influence on a recommendation required an explicit reference to an ERG's exploratory analysis or its ICER.

Results

Between September 2009 and September 2014, 100 STAs were undertaken by NICE that resulted in the production of guidance. In these 100 STAs, 40 different companies submitted documents as part of the NICE STA process. Full details of the disease areas covered, companies, and specific ERGs are reported in the article by Kaltenthaler et al. [2].

The Critique of the Submission

The ERG reports and their exploratory analyses

Most of the ERG reports (93%) for the STA process conducted and reported one or more exploratory analyses: seven ERG reports did not contain any exploratory analysis that generated a new ICER [8–14]. In the 93 reports that did include an exploratory analysis, the number of analyses ranged from 1 to 29, with an approximate mean of 8.5 analyses per report (798 analyses in 93 reports) and a median of 7 [2]. For the purposes of this research, the synthesis reduced these key data to whether an STA conducted more or less than the overall mean number of exploratory analyses. This approach permitted an analysis that made the most of the data without "stretching" it too far. For the 93 ERG reports that generated at least one exploratory analysis, the type of analysis that appeared in the largest proportion of reports was the category "Matters of judgment," that is, the ERG considered there to be uncertainty or possible variation regarding the evidence used to populate the model. At least one such exploratory analysis was conducted in 89% of the ERG reports. Some examples of text from ERG reports that are illustrative of "Matters of judgment" analyses are provided in Table 3 [15,16]. T3

Table 3 – Illustrative examples of the "Matters of judgment" category.

Quotation/extract
"In the MS the manufacturer uses the following utility values from the Oxford Outcomes study for PFS: PFS1 = 0.88 (disease free); PFS2 = 0.79 (remission/full response to therapy). The ERG considers that there should be no difference in the utility values used in the model to describe PFS1 and PFS2 as both groups of patients are in "remission/full response." When corrected by the ERG, this revision has the effect of reducing the QALY gain in PRIMA PFS by more than 10% and will therefore increase the ICER by approximately 11%." [15]
"In the manufacturer's base case it is assumed that the frequency of cardiac monitoring is greater when the patient is receiving epirubicin than when receiving trastuzumab. However, the ERG clinical advisors suggested that cardiac monitoring may not routinely be undertaken when treating patients with epirubicin. Consequently, the ERG undertook an additional sensitivity analysis which assumed that the monitoring frequency for epirubicin is equal to that for trastuzumab. This increased the ICER to È50,816 per QALY." [16]

ERG, evidence review group; ICER, incremental cost-effectiveness ratio; MS; PFS; QALY, quality-adjusted life-year; PRIMA. Q6

Table 4 – Illustrative examples of the "ERG base-case" category.

Quotation/extract
"To explore the potential impact of altering a range of separate assumptions simultaneously the ERG undertook two 'alternative base case' analyses which altered key assumptions of the manufacturer's model for which alternative estimates or assumptions were considered equally plausible to those employed by the manufacturer … The combined impact of making these three changes to the model changed the comparator to EOX and increased the ICER of HCX vs EOX to £66,982 per QALY."[16]
"New model results were generated by the ERG to take account of each of the issues previously identified … the separate effect of each change is shown in the upper section of Table 6-1 compared to the manufacturer's submitted base-case analysis. …The combined effect of these changes is to increase the incremental cost attributable to use of pemetrexed by 35% as well as reducing the incremental QALYs gained by 2%, so that the ICER increases from £33,732 to £47,239 per QALY gained." [17]

EOX; ERG, evidence review group; HCX; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life-year.

As a category of exploratory analysis, the calculation of a new base case by an ERG appeared in 48% of reports. For some text examples of this type of analysis, including the impact of these analyses on the base-case ICER [16,17], see Table 4.

Much less frequent were analyses that were concerned with either fixing unequivocal errors in the submitted models (35%) or addressing violations (18%). The total numbers for all types of analysis add up to more than 100% because an ERG report could include more than one type of analysis. Some illustrative examples of noted errors and violations in ERG reports are provided in Table 5 [15,17].

Table 5 – Illustrative examples of the "Fixing errors" and "Addressing violations" categories.

Quotation/extract
Fixing errors
"The submitted model uses data from the … trial to estimate the proportion of patients failing from the first PFS period … but who did not progress to second-line induction therapy. These proportions were calculated relative to the whole randomised population, but are applied in the model only to those patients still alive at the end of PFS. The ERG's correction of this anomaly has only a minor effect on the base-case estimates, increasing both incremental costs and outcomes, and raising the ICER by £19 per QALY gained." [15]
"A minor [arithmetic] error has been detected in calculating the proportion of patients assumed to receive docetaxel and erlotinib in second-line therapy. When this is corrected the ICER for the manufacturer's base case rises slightly to £33,817 per QALY gained." [17]
Addressing violations
"The manufacturer has failed to implement discounting correctly according to UK practice (i.e. applied annually). … This change increased the incremental overall discounted cost per patient by £736 (+3.9%) and incremental discounted QALYs per patient by 0.019 (+1.6%), resulting in the ICER increasing by £370 per QALY gained." [15]

ERG, evidence review group; ICER, incremental cost-effectiveness ratio; PFS; QALY, quality-adjusted life-year.

Table 6 – Proportions of types of decisions in ACDs and FADs.

Decision type	Number of reports
ACD (n = 81)	
Recommended	10 (13%)
Optimized	4 (5%)
Minded No	19 (23%)
No	48 (59%)
FAD (n = 100)	
Recommended	51 (51%)
Optimized	21 (21%)
No	28 (28%)

ACD, appraisal consultation document; FAD, final appraisal determination.

The ACs

The ACs' recommendations

Following the discussion of the ERG report and other factors, ACs produce their ACDs. In 19 STAs an FAD was issued without the previous publication of an ACD because the submission and appraisal dictated that further major work was not required. An interim ACD was, however, produced for most of the STAs (81 of 100) because a positive recommendation without restrictions was not forthcoming. In the 81 STAs with an ACD, the ACDs reported few initial positive recommendations (18%), but most of the FADs certainly did so (72%) (see Table 6). Over the 100 STAs as a whole, it is easy to see the dramatic change in recommendations between the decisions recorded in the ACD and the final recommendation reported in the FAD, suggesting that a process of scrutiny, consultation, negotiation, and specification is at work (see Table 6).

The ACs' preferred ICERs

A summary of the ICERs submitted by companies, and the relationship with the preferred ICERs of ACs, as outlined in the ACDs and FADs, is presented in Table 7.

It is apparent how many technologies appear to be cost-effective at a threshold of £20,000 per QALY gained in the original company submissions, and how few maintain that level of cost-effectiveness after appraisal and the production of the ACD. Most of the ACDs (72 of 81; 89%) recorded preferred or most plausible ICERs that were greater than £20,000 per QALY gained,

Table 7 – Differences between the base-case ICERs submitted by companies and the ICERs preferred by the ACDs.

Company base-case ICER to ACD-preferred ICER	Number of reports (n = 81)
No change (≤£20,000)	9 (11%)
From ≤£20,000 to >£20,000	4 (5%)
From ≤£20,000 to "no preferred" or "no plausible ICER"	20 (25%)
From ≥£20,000 to "no preferred" or "no plausible ICER"	13 (16%)
No change (>£20,000)	35 (43%)

ACD, appraisal consultation document; ICER, incremental cost-effectiveness ratio.

Table 8 – FAD ICERs.

FAD finding	Number of reports (n = 100)
With preferred ICER ≤£20,000*	37
With preferred ICER £20,000–£30,000	6
With preferred ICER >£30,000	36
AC did not specify a preferred ICER†	21

AC, Appraisal Committee; FAD, final appraisal determination; ICER, incremental cost-effectiveness ratio.

* Including ranges with at least one ICER less than £20,000 or an acknowledgment that the technology is cost-effective at the £20,000 threshold.

† Includes FADs without an ICER in which the analysis was based on "costs" alone.

or stated that there was no plausible ICER because of uncertainties within the evidence and model. By the production of the FAD, almost half of the STAs (43%) had achieved a preferred ICER of £20,000 or £30,000 per QALY gained, or less, and all 43 technologies received a positive recommendation (see Table 8). These included the19 STAs without an ACD, 15 (i.e., 79%) of which had a company's ICER less than £20,000 per QALY gained.

Nevertheless, we know that, by the production of the FAD, 72 STAs had received a positive recommendation, out of which 29 STAs received a positive recommendation with a preferred ICER in excess of £30,000 per QALY gained, underlining that this "threshold" is only a guide. In only 9 of the 72 STAs did the ACs conclude that the technology satisfied NICE's end-of-life criteria, which permit higher ICERs [2]. Nevertheless, another 12 STAs with high ICERs, and which took account of end-of-life criteria, failed to gain a positive recommendation.

The Influence of the ERGs' Exploratory Analyses

The ICERs preferred by ACs were often higher than the base-case ICER submitted by a company. This is principally a consequence of the exploratory analyses conducted by the ERG. It is not the case that these ERG analyses simply always generated a single, higher base-case ICER than that submitted by the company. In fact, the ERG analyses would most often generate a number of different ICERs, with the aim of testing the findings of the submission and providing useful information for the committees on the basis of different scenarios. A typical example can be found in the STA on rivaroxaban for the prevention of stroke and systemic embolism in atrial fibrillation [18]. The company submission reported a base-case ICER of £18,883 per QALY for rivaroxaban versus its principal comparator, warfarin. The ERG conducted 21exploratory analyses concerning the warfarin comparison, generating a range of new ERG-preferred base-case ICERs up to £62,568 per QALY gained. The conclusion of the first AC was that it did not consider the estimates of cost-effectiveness of rivaroxaban compared with warfarin from the present model to be appropriate, but accepted the ERG base case that the ICER was likely to be up to £62,568 per QALY. The final decision, after further analyses, was that the most plausible ICER was between £2,870 (the company's post-ACD analyses) and £29,500 per QALY gained, and so the ICER was within the range that could be considered a cost-effective use of NHS resources. The company's ICER was therefore relevant, and exploratory analyses generated the highest ICER, but the final decision was made on the basis of consideration of a range of ICERs.

Table 9 presents the source of the ICERs preferred by the ACs, as stated in the ACDs and FADs from the STAs included in this analysis. Often, more than one preferred ICER was presented in the documents and the source of the preferred ICER was not always clear in the ACD and FAD. Of the 81 STAs with ACDs, in most cases there was either no preferred ICER (38%) or the ICERs presented by the ERG were preferred by the AC (36%). More often than not, however, the most preferred ICERs included estimates from both the company and the ERG.

The ACs' recommendations

On the basis of the explicit referencing of exploratory analyses or their ICERs in the text of the ACDs and FADs that outlined the committees' decisions, the work of the ERG had a clear influence on recommendations (see Table 10).

The original ERG exploratory analyses, presented in the ERG reports, appear to have influenced recommendations in a smaller proportion of FADs (47%) than ACDs (72%). This can be explained by the work conducted between the ACD and the FAD: the ERG exploratory analyses influence the ACD, which impels the company to conduct more work, including rerunning analyses and generating new ICERs. It was rare for the ACD to mention a specific request for an ERG to conduct further analyses or critical appraisal, and when this did occur, it usually took place before the ACD. For example, it was stated in an addendum to one report that the ERG had been "requested to provide additional analyses for the STA of aripiprazole for the treatment of schizophrenia in adolescents (aged 15–17 years) … Can an indication of the cost-effectiveness of the first-line aripiprazole strategy compared with a first-line risperidone strategy be provided using the estimated costs for risperidone (for adolescent schizophrenia) and the manufacturer's economic model?" [19]. Nevertheless, it was most often the case for there to be no specific request made to the ERG at all, but rather a company would be required to conduct additional work in response to a negative recommendation in the ACD, and this work would then be critiqued further by the ERG. Sixty of the 81 STAs with an ACD (71%) involved additional work being performed by the ERG between the production of the ACD and the FAD. Of these, 97% (58 of 60) had received a "No" or "Minded No" recommendation in the ACD, with the ACD-preferred ICER being classified as either "unspecified" (20 of 60; 33%) or £20,000 or more per QALY gained (38 of 60; 63%). Finally, inevitably, another factor that influenced final recommendations was the submission by a company of a patient access scheme (PAS). A PAS was submitted after the production of the ACD in 23 of these STAs and led to a positive change in recommendation in 65% (15 of 23) of cases.

Table 9 – Source of preferred ICERs for ACDs and FADs.

	ERG	Company	Both ERG and company	No preferred ICER	Unclear
ACD (n = 81)	29 (36%)	9 (11%)	7 (9%)	31 (38%)	5 (6%)
FAD (n = 100)	27 (27%)	23 (23%)	17 (17%)	24 (24%)	9 (9%)

ACD, appraisal consultation document; ERG, evidence review group; FAD, final appraisal determination; ICER, incremental cost-effectiveness ratio.

Q10

Table 10 – Influence of exploratory analyses on NICE recommendations.

	Number of reports
ACD	
ACD mentions one or more exploratory analyses	74 of 76 (97%)
ACD recommendation is clearly influenced by one or more of the exploratory analyses mentioned	55 of 76 (72%)
FAD	
FAD mentions one or more exploratory analyses	87 of 93 (94%)
FAD recommendation is clearly influenced by one or more of the exploratory analyses mentioned	44 of 93 (47%)

ACD, appraisal consultation document; FAD, final appraisal determination; NICE, National Institute for Health and Care Excellence.

Discussion

As demonstrated by the frequency, number, and type of analyses performed, ERG exploratory analyses were not conducted simply to generate a single, alternative ICER, but rather to present a number of corrections and scenario analyses, presumably with the intention of allowing an AC to decide which ICER it preferred as being the most appropriate or relevant. The aim was to ensure that the AC had sufficient robust information to reach a decision. When it was deemed that this had not been provided by the company, the ERG was responsible for undertaking what they determined to be the most appropriate exploratory analyses. ERGs not only test the company model but also aim to anticipate what information the AC will need to make its decisions. At least one of the exploratory analyses conducted and reported by an ERG is mentioned in 97% of the ACDs and had a clear influence on more than two-third of ACD recommendations. The influence of these original exploratory analyses appears reduced in the FAD, although they still had some direct influence on the final recommendations in almost half of the STAs. On the whole, the influence of the original exploratory analyses becomes more indirect: they inform the AC's preferred assumptions, which often form the starting point for further company analyses undertaken after the ACD is published. Consequently, it is apparent that the ERG exploratory analyses are highly influential in the ACD but are necessarily superseded by the later, related, additional work of companies and ERGs, after the ACD, including when a PAS has been submitted, which effectively lowers the price of the technology [2].

It cannot be assumed that the number of exploratory analyses is an indicator of the quality of a company submission. As stated elsewhere with reference to STAs with no such analyses [2], this absence might be due to an ERG concluding that a model is either of very good quality or of very poor quality. If a model is considered to be of very poor quality, the ERG might decide that undertaking additional exploratory analyses would add no value to the decision-making process. Nevertheless, the type of exploratory analysis undertaken certainly might function as a guide to the quality of the economic model in a company submission. If the ERG is performing analyses to fix errors or address violations, then this suggests that there were unequivocal issues with the perceived quality of the submitted model. In this sample, ERGs conducted at least one such analysis, respectively, in 35% and 18% of STAs. "Matters of judgment," however, was the principal type of analysis found in this sample of STAs: 89% of ERG reports with analyses contained at least one such analysis. ERGs therefore frequently saw value in exploring multiple alternative scenarios.

Research has found that there are similarities in the economic evaluation processes conducted for health technology assessment (HTA) in different countries, especially in Europe [20], but only one other recent study has conducted a critique of cost-effectiveness analyses submitted by companies to an HTA process. The study by Ghabri et al. [3,21] has explored the critical opinions of the Economic and Public Health Evaluation Committee regarding uncertainties in the submitted cost-effectiveness analyses of the first 28 submissions to the French National Authority for Health. This study classified cost-effectiveness uncertainty using standard international (International Society for Pharmacoeconomics and Outcomes Research) and HTA frameworks and identified issues similar to the present study. Ghabri et al. identified methodological, parameter, and structural uncertainties, which map in a general way to many of the issues being addressed by this article's exploratory analysis categories (e.g., the methodological uncertainties in the study by Ghabri et al. included adherence to the reference case and required guidelines, similar to the present study's "Addressing violations" category). Previously, Hill et al. [22] had also identified similar problems with the models considered by the Australian Pharmaceutical Benefits Advisory Committee (e.g., questions over technical aspects of the model, unsubstantiated assumptions, and calculation errors). These studies have therefore also considered the limitations of submitted economic models, but the present study goes further, detailing explicitly how such limitations are addressed within the NICE process and how far these actions influence decision making. The findings of this study therefore have relevance to the conduct and critique of cost-effectiveness analyses internationally.

The strengths of this research are that this was an analysis of the most recent 100 STAs, which offers a good summary of present and recent practice. The development of a simple coding scheme, the extensive piloting of the data extraction tool, and the double-checking of all key data across the 100 STAs by at least two experienced health economic modelers reduced the likelihood of inconsistency and inaccuracy in the data. In addition, the method of synthesis was principally descriptive, which reduced the likelihood of overstating relationships in the data, and a reductive approach was taken to managing data that might be affected by interpretation or by poor reporting in the original documents.

There are, however, some limitations in this study. There are inherent weaknesses in using documentary analysis in that the researcher is able to analyze only what has been reported. The level and type of detail provided in and across the ERG reports and other documents could be very different, which made data extraction time-consuming, difficult, and at times a matter of interpretation. The description of the analyses undertaken was often highly specific to a particular STA and could be inconsistent across ERG reports. This could make coding difficult. The source of ICERs cited in ACDs and FADs could also be unclear or open to interpretation. Despite efforts to simplify the data extraction and coding process, it was not possible to report reliably or precisely, for example, what exact proportion of the 100s of exploratory analyses was actually mentioned in ACDs or FADs, or influenced their recommendations, and so this has been reduced simply to whether "one or more" exploratory analyses were mentioned or had an influence in any particular STA. In addition, the data did not permit a deeper exploration of the nuances and complexities within the analyses, for example, the types of models being used or the implementation of the exploratory analyses, including the sources of data [2].

Data extraction was undertaken by two research teams (Sheffield and York) that between them had undertaken nearly a third of the STAs. Although this introduced a potential source of bias, this was minimized to some extent by having neither team extracting data from its own reports. In addition, the inside knowledge obtained from working on so many STAs was crucially important in interpreting the data. These limitations suggest that caution should be exercised regarding some conclusions drawn from the evidence, especially concerning the generalizability of any findings.

Additional future research priorities include a prospective qualitative study of a limited number of STAs to assess how ERGs make decisions regarding which exploratory analyses should be undertaken, and to determine which type of exploratory analyses AC members find most useful when making recommendations. A study to explore how the presence and extent of the exploratory analyses might vary according to the skills, experience, and judgments of the ERGs would also be useful, as would an in-depth analysis of the category of "Matters of judgment." This could be done by prospectively categorizing the nature of the implementation of any exploratory analysis and the data sources used, for example, whether an analysis was based on different but equally valid assumptions or different but equally valid sources of data. An assessment of the impact of unequivocal errors on company ICERs would also be worthwhile.

Conclusions

Exploratory analyses are undertaken by the ERGs as part of the NICE STA process to help inform decision making. They are intended to provide support to ACs by addressing errors in company submissions and conducting scenario analyses to reduce the uncertainties surrounding a technology within a particular health service context. The results of this study suggest that the additional analyses undertaken by ERGs in the appraisal of company submissions are highly influential in the policymaking and decision-making process.

Source of financial support: This project was funded by the National Institute for Health Research HTA Programme. For more details, visit http://www.nets.nihr.ac.uk/projects/hta/1415104. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HTA Programme, NICE, the National Institute for Health Research, the NHS, or the Department of Health.

REFERENCES

[1] National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal 2013. London: National Institute for Health and Care Excellence, 2013.
[2] Kaltenthaler E, Carroll C, Hill-McManus D, et al. The use of exploratory analyses within the National Institute for Health and Care Excellence single technology appraisal process: an evaluation and qualitative analysis. Health Technol Assess 2016;20:1–48.
[3] Ghabri S, Hamers FF, Josselin JM. Exploring uncertainty in economic evaluations of drugs and medical devices in France. Pharmacoeconomics 2016;34:617–24.
[4] Tappenden P, Chilcott JB. Avoiding and identifying errors and other threats to the credibility of health economic models. Pharmacoeconomics 2014;32:967–79.
[5] Ritchie J, Lewis J. Qualitative research practice: a guide for social science students and researchers. London: Sage, 2003.
[6] Pope C, Mays N, Popay J. Synthesizing Qualitative and Quantitative Health Evidence: A Guide to Methods. Maidenhead, UK: Open University Press, 2007.
[7] McCabe C, Claxton KF, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. Pharmacoeconomics 2008;26:733–44.
[8] National Institute for Health and Care Excellence. Azacitidine for the Treatment of Myelodysplastic Syndromes, Chronic Myelomonocytic Leukaemia and Acute Myeloid Leukaemia. NICE Technology Appraisal Guidance TA218. London: National Institute for Health and Care Excellence, 2011.
[9] National Institute for Health and Care Excellence. Certolizumab Pegol for the Treatment of Rheumatoid Arthritis. NICE Technology Appraisal Guidance TA186. London: National Institute for Health and Care Excellence, 2010.
[10] National Institute for Health and Care Excellence. Ivabradine for Treating Chronic Heart Failure. NICE Technology Appraisal Guidance TA267. London: National Institute for Health and Care Excellence, 2012.
[11] National Institute for Health and Care Excellence. Afatinib for Treating Epidermal Growth Factor Receptor Mutation-Positive Locally Advanced or Metastatic Non-Small-Cell Lung Cancer. NICE Technology Appraisal Guidance TA310. London: National Institute for Health and Care Excellence, 2014.
[12] National Institute for Health and Care Excellence. Dapagliflozin in Combination Therapy for Treating Type 2 Diabetes. NICE Technology Appraisal Guidance TA288. London: National Institute for Health and Care Excellence, 2013.
[13] National Institute for Health and Care Excellence. Capecitabine for the Treatment of Advanced Gastric Cancer. NICE Technology Appraisal Guidance TA191. London: National Institute for Health and Care Excellence, 2010.
[14] National Institute for Health and Care Excellence. Pemetrexed for the First-Line Treatment of Non-Small-Cell Lung Cancer. NICE Technology Appraisal Guidance TA181. London: National Institute for Health and Care Excellence, 2009.
[15] National Institute for Health and Care Excellence. Rituximab for the First-Line Maintenance Treatment of Follicular Non-Hodgkin's Lymphoma. NICE Technology Appraisal Guidance TA226. London: National Institute for Health and Care Excellence, 2011.
[16] National Institute for Health and Care Excellence. Trastuzumab for the Treatment of HER2-Positive Metastatic Gastric Cancer. NICE Technology Appraisal Guidance TA208. London: National Institute for Health and Care Excellence, 2010.
[17] National Institute for Health and Care Excellence. Pemetrexed for the Maintenance Treatment of Non-Small-Cell Lung Cancer. NICE Technology Appraisal Guidance TA190. London: National Institute for Health and Care Excellence, 2010.
[18] National Institute for Health and Care Excellence. Rivaroxaban for the Prevention of Stroke and Systemic Embolism in People with Atrial Fibrillation. NICE Technology Appraisal Guidance TA256. London: National Institute for Health and Care Excellence, 2012.
[19] National Institute for Health and Care Excellence. Aripiprazole for the Treatment of Schizophrenia in People Aged 15 to 17 Years. NICE Technology Appraisal Guidance TA213. London: National Institute for Health and Care Excellence, 2011.
[20] EUnetHTA Joint Action 2 WP7S3; Heintz E, Gerber-Grote A, Ghabri S, et al. Is there a European view on health economic evaluations? Results from a synposis of methodological guidelines used in the EUnetHTA partner countries. Pharmacoeconomics 2016;34:59–76.
[21] Ghabri S, Hamers FF, Josselin JM, Harousseau JL. Exploring uncertainty in economic evaluation of medicines: a review of the first manufacurers' submissions to the French National Authority for Health (HAS). Value Health 2014;17:A441–2.
[22] Hill S, Mitchell A, Henry D. Problems with the interpretation of pharmacoeconomic analyses: a review of submissions to the Australian Pharamceutical Benefits Scheme (PBS). JAMA 2000;283:2116–21.