PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	How long has NICE taken to produce Technology Appraisal guidance? A retrospective study to estimate predictors of time to guidance.
AUTHORS	Casson, Steven; Ruiz, Francis; Miners, Alec

VERSION 1 - REVIEW

REVIEWER	Priv.Doz.Dr.phil. Claudia Wild Institutsleiterin/ Director Ludwig Boltzmann Institut/e für/for Health Technology Assessment Garnisongasse 7/20 1090 Wien/ Vienna Austria
REVIEW RETURNED	27-Aug-2012

REPORTING & ETHICS	The only question I would like to answer is that of relevancy: I can not find this article adding information to the already known.
GENERAL COMMENTS	The authors of the article intend to analyse the time needed within NICE for producing STAs and to identify explanatory variables for predicting faster or delayed guidance. In contrast to other authors (Ford and to ONeill) attempting to analyse a similar question, Casson et al. claim to measure only the time that is under control from NICE, namely the time between the DoH´s request for an assessment and the actual publication of the guidance. With focusing on this time-period only (additionally limited with measuring the time from start of appraisal excluding the scoping period) the authors are answering solely a question that is of limited relevance. Of interest for NICE` internal managerial recipients and not for a broader audience, esp. those criticising (manufacturers and patient advocates, general public) the slow processes from product launch to guidance. Besides the limited relevancy of the question seeking for an answer the methodology chosen „survival analysis/ time to event“ is rather sophisticated and questionable if a qualitative analysis of the variables influencing the lengthiness of the process would not have added more in-depths information than a quantitative analysis. After this „survival analysis/ time to event“ we know, that the STA process has resulted in faster guidance (compared to MTA) and appeals delay the process considerably and that cancer-related technologies are more often appealed to. None of those informations are new, no relevant additional information is added to the Ford and O`Neill publication.

	Qualitative information on approval regime (orphan designation, under exceptional circumstances) or existance and role of patient advocat groups in appeals etc. as variables influencing the lengthiness of the process - would add substantial pieces of information. Comments in detail: • Survical analysis should – for the sake of better understanding – be called „time to event“ (self-explanatory) analysis all through the document, esp. in the abstract. • That time to guidance has nothing to do with „quality of decision“ is self-evident and does not have to be stated. It rather confuses, if there was supposed to be information on „quality“ of decision. • …the institute was asked to assess the cost-effectiveness of „faller´s clinics“……irrelevant information • 13th February is a rather arbitrary end date. Why not end of Jan or end of Feb ? is there an explanation ? • There are no/ hardly any commas all through the text. I am not a native speaker, but in my language understanding, commas are lacking here. • P 9: „ultimately“ is used twice within the paragraph….finally, at end, eventually….

REVIEWER	Stirling Bryan Professor, School of Population & Public Health University of British Columbia Director, Centre for Clinical Epidemiology & Evaluation Vancouver Coastal Health Research Institute No competing interests
REVIEW RETURNED	07-Sep-2012

THE STUDY	My main concern is that I see the question as being flawed! By definition STAs and MTAs are different - one focuses on single technologies and the other on multiple technologies! And so why is it interesting to compare the time taken to undertake two rather different tasks? Isn't is obviously that a more complex task, where multiple technologies are being assessed, will be a longer process?
RESULTS & CONCLUSIONS	I think the question needs to be more fully justified, in light of my comments above. This will also require some strengthening of the discussion of study weaknesses. I would also have liked to see some more exploratory data analysis, looking graphically at relationships between variables available for analysis. Did the authors consider the scale of the evidence base as a variable? I would hypothesise that where there exists only a small number of trials, the time would be shorter.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1(Wild)

1. The reviewer questions the novelty of the research, they state that the information is already known. We accept that this is not the first time this subject has been studied, and we clearly state this and reference the Barnham / Ford / O'Neill studies in the paper. However, we maintain that this study is different because we define the start of an appraisal differently(and we believe more appropriately), we included studies that haven't been completed (we treat them as censored events) and we use formal statistical techniques to adjust for censored events. Therefore, while it can be argued that the aims were similar to those from previous studies, the methods are quite different. Indeed O'Neill et al recommended that their observational trend findings should be revisited using more formal statistical approaches. Moreover, we should add that while it might be known that factors such as STA (yes / no) result in shorter time to guidance, the extent of this time saving has not been clearly stated, and has not been previously analysed using multivariate techniques. That is, to isolate the independent effects of the predictor variables. For these reasons, we disagree with the general sentiment expressed by the reviewer.

2. The reviewer questions the relevance of the time period we have chosen to analyse, compared to previous studies. We have analysed the time from NICE being asked to appraise a technology to the time of publication of guidance. Other studies have used various other starting points (e.g. time from product launch). The choice of guidance publication as end date has not been questioned. The reviewer states that our choice of start date means that the results are only of interest to 'NICE's internal managerial recipients'. We are sure that our analysis will be of interest to this group; however we believe all other consultee groups (manufacturers, patient, clinical groups - those who submit to NICEs Appraisals Programme) will also be interested. This part of the guidance production process is still a crucial component even if it does only relate to part of its sum. Moreover, we maintain that it makes little sense in terms of understanding NICE's role in producing guidance to the NHS if longer times are included. For example, we give an example (the vinorelbine appraisal) of when the time of product launch was used as the analysis start date. This was 2 years before NICE existed which is clearly nonsensical. Therefore, even if the analysis is someway 'restricted', we believe that using the scope date as the beginning of the appraisal time is more logical / more of a fair comparison in terms of assessing NICE's role in the production of guidance. Including earlier start times could easily lead to erroneous results.

3. The reviewer questions the need to use 'sophisticated' survival analysis techniques. Qualitative and simpler quantitative approaches certainly have a role, but they are limited. As already mentioned (see reply no. 1) the need to use more advanced statistical approaches (in order to formally identify independent effects) was one of O'Neill's conclusions. Moreover, using this approach we are able to include appraisals that have not completed (and are potentially very long) as they can be treated as censored events. Lastly, because we used this approach, we were able to generate quantitative estimates of the independent effects of each predictor variable (e.g., appeals, STA process etc). Therefore while we accept it is known that factors such as STA process (yes / no) and appeals (yes /no) speed up and delay processes respectively, we are able to provide estimates of how much time they add. We note that the reviewer states that it is known that cancer-related appraisals are more often appealed against. We are unsure which analysis the reviewer is referring to; it is not a conclusion we have drawn.

4. We agree with the reviewer that our analysis has not / is unlikely to have captured all predictors of the time to guidance, and there are others such as those the reviewer lists. We also agree that this is where qualitative research would help. However, for the record, and with relevance to our analysis, many of these variables were considered for inclusion in the original analysis but there wasn't and often isn't a robust way of operationalising these variables in a quantitative sense. For example, patient groups are always involved in appeals, in so much as they have a right to appeal and there is often more than one group, sometimes providing joint submissions. Moreover, in STAs, only the product manufacturer has a right to appeal. In recognition of this, we have added the following text to the discussion "although it is likely that other variables may be related to the time to guidance, there

are often challenges in quantifying them. One such example is the number or mix of consultees, which could reflect the complexity / level of interest in a particular area…" ".
5. Regarding the methods. We have used time to event methods to address all appraisals regardless of completion status, and also multivariate analysis. Our finding that the evidence of durational increase over time is weak adds to the literature knowledge on the subject. Other authors, including Ford et al (2012) and Kaltenthaler et al (Health Technology Assessment 2011; Vol. 15: No. 22), have addressed qualitative aspects related to the technology appraisal processes and it was not our intention to address such qualitative issues in any detail other than to make passing reference.
6. We have changed the term 'survival analysis' to 'time to event analysis as requested.
7. We have removed the statement to do with the 'quality of the decision' as requested.
8. The information referring to the 'faller's clinic' has been deleted.
9. It has been clarified that 13th February related to the end of the data collection period. The date in itself has no particular significance; rather it related to our project timetable.
10. The reviewer states that more commas are needed through the script. We have thoroughly reread the document and edited the grammar where we believe it to be appropriate. The repetition of the word 'ultimately' had been addressed.
Reviewer 2 (Stirling)

11. The reviewer questions one of the design elements of the research project, that being a comparison of MTAs and STAs when they are different tasks. We strongly disagree with this assertion, for a number of reasons. First, the type of process (STA or MTA) is a covariate in the analysis; the main question was to examine the time to produce guidance. It is not a comparison between MTA and STA processes per se. Second, it is a valid comparison since both are designed to produce the same objective (i.e. guidance to the NHS). They are largely identical in structure (but not scheduling) with the exception of the sub-process which assesses the evidence of effectiveness and cost-effectiveness. Third, previous authors in the peer-reviewed literature have also felt this to be a valid comparison e.g. Ford et al (2012), which is published in BMJ Open. We respectfully do not feel therefore, that there is a need to provide further justification for the question.
12. While we believe the study to be an improvement on existing research, we agree that there are a number of weaknesses / limitations with it. A number of these limitations are already in the discussion, including some of the difficulties with comparing STA and MTA processed, but we have added to them following the comments from Reviewer 1. Namely, that we are unlikely to have included all independent variables of interest, and that qualitative research would be useful.
13. The reviewer asks for further exploratory analysis and related graphs, presumably to other potential predictors of the time to guidance. We have collected data on a number of other variables. However, many variables, such as the number of consultees, are difficult to include in a statistical analysis because they are correlated with other variables (e.g. officially there is only one consultee in a STA process). Moreover, there is little prior reason to believe that others (e.g. type of disease) are likely to predict results, and numbers are too low to produce any meaningful results, even if simply graphed. The reviewers next point, 'did we consider the scale of the evidence base', exactly highlights some of these problem. While we agree with the logic that more complex / detailed evidence bases are likely to take longer to conclude, how should this be defined and consistently applied to all appraisals? By the number of RCTs, economic evaluations, submissions, responses to consultation? What is a relevant RCT? It is unclear. We would therefore respectfully prefer not to do as the reviewer requests, but at the same time we are happy to leave this to an editorial decision.

VERSION 2 – REVIEW

REVIEWER	Stirling Bryan Professor University of British Columbia Canada
REVIEW RETURNED	29-Nov-2012

GENERAL COMMENTS	I am concerned that the sole focus for the paper is time. I don't deny the importance of timely guidance and there is no doubt that the STA is quicker but surely we need also to consider the quality of the process - speed is surely not everything. The omission of independent analysis in the STA cannot increase quality and so the question is whether, and to what extent, the quality of analyses has been compromised. Previous evidence from Miners suggests that industry and academic analyses give quite different results - this would seem to be an important discussion issue that is largely ignored by the authors.

VERSION 2 – AUTHOR RESPONSE

The sole comment (from Professor Stirling) related to the quality of the STA process. In response, we have added an additional paragraph (with an associated reference) to the discussion section which raises the issue of quality of evidence submissions, and suggests, from the authors' perspectives, possible shortcomings of the STA process. At the same time we make it clear that our analysis says nothing about the quality of submissions.

Additionally, we have amended one typographical error (see 'Strengths and limitations' section), amended three other sentences to improve the semantics of the text (see 'Funding', 'How does this compare to other studies', and 'Strengths and limitations' sections), and added a statement that the views expressed are those of the authors alone, and not necessarily those of any associated organisation, e.g. NICE (see 'Declaration of Interests' section). All changes are clearly highlighted in 'tracked-changes' (red or dark blue type-face).

We are very encouraged to hear that the reviewers have recommended publication and we hope the included revisions will be sufficient for publication to proceed.