

BMJ Open Involving clinical experts in prioritising topics for health technology assessment: a randomised controlled trial

Andrew Cook,^{1,2} Elke Streit,³ Gill Davage³

To cite: Cook A, Streit E, Davage G. Involving clinical experts in prioritising topics for health technology assessment: a randomised controlled trial. *BMJ Open* 2017;7:e016104. doi:10.1136/bmjopen-2017-016104

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmj-2017-016104>).

Received 27 January 2017

Revised 30 May 2017

Accepted 23 June 2017

ABSTRACT

Objectives The objective of this study was to explore whether reducing the material supplied to external experts during peer review and decreasing the burden of response would maintain review quality into prioritising research questions for a major research funder.

Methods and analysis Clinical experts who agreed to review documents outlining research for potential commissioning were screened for eligibility and randomised in a factorial design to two types of review materials (long document versus short document) and response modes (structured review form versus free text email response). Previous and current members of the funder's programme groups were excluded. Response quality was assessed by use of a four-point scoring tool and analysed by intention to treat.

Results 554 consecutive experts were screened for eligibility and 460 were randomised (232 and 228 to long document or short document, respectively; 230 each to structured response or free text). 356 participants provided reviews, 90 did not respond and 14 were excluded after randomisation as not eligible. The pooled mean quality score was 2.4 (SD=0.95). The short document scored 0.037 (Cohen's $d=0.039$) extra quality points over the long document arm, and the structured response scored 0.335 (Cohen's $d=0.353$) over free text. The allocation did not appear to have any effect on the experts' willingness to engage with the task.

Conclusions Neither providing a short or a long document outlining suggested research was shown to be superior. However, providing a structured form to guide the expert response provided more useful information than allowing free text. The funder should continue to use a structured form to gather responses. It would be acceptable to provide shorter documents to reviewers, if there were reasons to do so.

Trial registration number ANZCTR12614000167662.

INTRODUCTION

Chalmers and Glasziou have suggested that as much as 85% of the US\$100 billion spent on health research worldwide each year is potentially wasted due to four key problems of knowledge production and dissemination. These four areas include (1) ensuring that the right research questions are asked; (2) ensuring that study designs are appropriate and are of methodological

Strengths and limitations of this study

- The trial included all eligible clinical experts over the course of a year.
- The largest effects were shown in areas where assessors could not be masked. The lack of ability to blind assessors to one of the two allocations is a weakness.
- The findings will directly influence practice in a major clinical trials funder.

quality; (3) ensuring that the findings from funded research are available in the public domain; and (4) ensuring that funded research is unbiased and usable.¹

The National Institute for Health Research (NIHR) Health Technology Appraisal (HTA) programme was established in the 1990s, in part to address market failure in UK health research, and is now embedded in NIHR, managed by the NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC). The programme is the major public funder of pragmatic trials in the UK, and its range of activities are discussed elsewhere.^{2,3}

In the commissioned mode, the HTA programme decides on the research question to be answered in the light of National Health Service (NHS) need, and it advertises commissioning briefs for teams of researchers to bid competitively for funding to deliver the answers. The prioritisation and refinement of the question within the commissioned mode is one of the key ways in which the programme can interact with NHS clinicians and other stakeholders to ensure that it is asking the right questions — those to which the NHS needs answers.

The main tool which the HTA programme uses in commissioned mode for prioritising and refining research questions is the Topic Identification and Development (TIDE) panel. These are standing groups of up to 20 clinicians and lay members, grouped by



CrossMark

¹Wessex Institute, University of Southampton, Southampton, UK

²University Hospital Southampton NHS Foundation Trust, Southampton, UK

³NIHR Evaluation, Trials and Studies Coordinating Centre, University of Southampton, Southampton, UK

Correspondence to

Dr Andrew Cook;
andrewc@soton.ac.uk

clinical theme. The exact configuration of the panels varies over time. The current list can be found on the programme's website.⁴

Currently, the programme has five TIDE panels with approximately 20 members each — so it would be impossible for all appropriate expertise to be represented within a panel. Therefore, external clinical experts are used to inform and challenge each panel's opinions in much the same way that referees or peer reviewers are used by research funding boards. The programme secretariat prepare a vignette (a paper of four to eight pages, summarising the clinical dilemma, existing research and research under way) to inform the panel's discussion.

Under the established process, clinical experts are asked to comment on the vignette. They are approached with an email inviting them to contribute, and they are warned that the required work may take about an hour. If they accept, they are then sent the vignette and a structured form to complete and return to the secretariat. The secretariat then either update the vignette or pass the comments onto the TIDE panel for consideration. Sometimes the secretariat will iterate a point with the clinical expert.

Around 30% of experts approached will accept the offer to contribute to the programme. There are two related concerns about this low figure. The first is that the validity of the programme's approach to answering NHS relevant questions depends on interaction with the NHS. The second is that this rate of response may introduce bias—in that clinicians with particular opinions may be more likely to respond to invitations to participate. The combination would mean that the programme's outputs are not representative of NHS need. One way of addressing this would be to improve clinician participation—but not at the cost of the quality of advice received.

While there is a literature on peer review for the assessment of research applications and scientific papers,^{5–10} the literature on how to engage clinicians (not necessarily academics) in the prioritisation of research questions is sparse. We were unable to find anything of direct relevance to the HTA programme, so we had to consider what evidence we needed in order to refine the processes which we use to develop the research questions that we address to inform UK NHS practice.

An alternative model for engaging clinicians at this stage had been identified in discussion between the secretariat and two new TIDE panel chairs. In this model, clinicians would be asked to comment on the commissioning brief—a document of less than a page in length which summarises the research question to be asked, but not the background information. It was felt safe to assume that expert clinicians would be up to date with developments in their field. With a shorter document to consider, it was felt that the time for the work could be specified as 5–10 min, and rather than asking respondents to complete a form, the programme would accept responses as a reply to the initial invitation email.

We hypothesised that all these alterations to the process would serve to reduce friction and increase participation.

Objectives

We set out to investigate whether reducing the material supplied to external experts and decreasing the burden of response could be done without decreasing the usefulness of the input they provide. We were also interested in whether decreasing the burden of engaging with the programme would lead to increased participation (i.e., a greater proportion of experts accepting the invitation to participate and returning a useful response) and whether the method of identifying a potential expert was related to their willingness to contribute to the programme.

Methods

We conducted a factorial randomised controlled trial. One randomisation was between receiving a vignette and a commissioning brief. The other one was between being asked to respond using free text and being sent a structured form to complete.

Trial registration

We sought to register this trial prospectively with several trial registries. All declined to register it on the ground that no patients or measurable patient outcomes were involved. As registration seemed a remote possibility, and as the trial was intended to influence our own practice, we started the trial regardless.

About a month after recruitment started, we identified a paper¹¹ reporting a trial evaluating training for medical students, and we noted that it had been registered with the Australia and New Zealand Trial Registry. We therefore contacted that registry, which agreed to register our trial retrospectively, about 2 months into our 1-year recruitment period.

Participants and sample size

The participants were all clinical experts approached to comment on HTA commissioned mode research topics in 2014. This was selected as a pragmatic sample — the programme was willing to adapt its procedures to accommodate the study for up to one year. Over the course of the year, clinical experts agreed to comment on possible research on 554 occasions, and of these, 460 were randomised, the others being ineligible for the trial.

For experts approached to contribute to more than one vignette during the recruitment period, only their involvement with the first vignette was included in the study. This was to avoid clustering effects from including the same expert multiple times and also to avoid exposing individuals to multiple interventions. Experts were also excluded if they were current or previous members of HTA programme groups — such as the TIDE panels or funding boards — or if they had been consulted as methodology experts or as members of the public.

Randomisation and masking

Randomisation was conducted using a computer-generated sequence of permuted blocks of sizes 2, 4 and 6 in a 1:1 ratio. Each randomisation had its own block list, kept by the trial manager. When a new participant presented, they were assigned the next available allocation from each list. In the event of more than one participant being available for randomisation, they were ordered by the time their acceptance to participate email was received and the earlier acceptance allocated first.

Participants were informed that a research project was under way, but they were not informed of the hypothesis being tested as we believed that this knowledge would be likely to affect responses received. This was discussed and agreed with the University of Southampton Faculty of Medicine Ethics Committee.

HTA staff assessing the responses received were aware of the hypotheses being tested, but they were not informed of the allocation of participants who provided the responses that they were assessing. However, whether the response was provided as a free form text or in a structured form was simple for assessors to guess.

Outcome measures

The primary outcome was the usefulness of responses received, as measured by a quality score (from 0 - no review returned, to 4 - very helpful review) applied by the team responsible for preparing the vignette. As we did not know the behaviour of this score, we decided prospectively that superiority by a Cohen's *d* of 0.3 indicated a worthwhile effect which the programme may choose to act on.¹²

We also set out to explore the relationships between

- ▶ allocation and likelihood of responding
- ▶ the source of identification of the expert and the likelihood of responding.

We planned to assess the quality of masking by investigating the assessors ability to identify the document allocation.

Statistical methods

The usefulness of response was assessed by intention to treat, by assigning non-response a score of 0 (as not contributing any information was judged to be of no value). Usefulness of the responses was modelled with the analysis of variance, with the quality being the response variable and the two allocations (vignette versus commissioning brief and free text versus form) as the input variables. Interaction was investigated.

For assessment of masking, *p* values were calculated using a binomial test, assuming that the correct guess rate would be 0.5 if masking were perfect.

The relationship between allocation and the likelihood of an expert return his work was explored using a test for equality of proportions.

The influence of the source from where the expert was identified on the likelihood of response was investigated using χ^2 tests.

All analyses were conducted with R.¹³

Sources of data

Data on vignette allocation and quality of responses received were collected specifically for this study. Data on expert's willingness to participate in the reviewing process were extracted from data routinely collected within the HTA programme for business purposes.

Internal feasibility phase

We established a set of stopping rules, to be tested after around one third of the primary outcome data points had been collected. This was to protect against any of the options being so bad as to undermine the prioritisation processes of the programme and to ensure that the trial processes could be run within the HTA programme.

The rules were to stop if

- ▶ experts could not be randomised in a robust manner or
- ▶ the quality scores returned by the assessors were overall lower than what would have been expected if our usual processes had been followed.

In addition, all incoming comments were reviewed by the trial manager and informally assessed for usefulness compared with comments received outside the trial.

Changes during the study

We changed the main outcome measure early on in the study. Initially we asked assessors to score the usefulness of an expert response on a scale of 0–10. After the first 10 or so responses had been scored, there was a general view from the assessors that the scale was generally too detailed, and a one-point difference in the scale was not well understood. We revised the scale to 0–4 and asked our assessors to rescore the initial set of responses, and the assessors found this much more satisfactory. Under both systems, assessors were not allowed to express fractional values.

We modified the inclusion criteria twice during the course of the study, to make them more restrictive.

First, we had to refine our definition of a clinical expert (as opposed to a methodological expert). This was precipitated by being challenged to randomise a statistician with considerable experience of the clinical condition discussed in the document he was asked to comment on. We took the view that we only wanted people with specific clinical experience, and we updated the inclusion criteria to make this clear.

Second, we were presented with a clinical expert who had already taken part in the study and were asked whether he should receive the same allocation or be rerandomised. We took the view that, if rerandomised, part of the study hypothesis would likely be revealed to the expert and possibly influence their submission, and in any case, it was likely that the scoring for all responses from an individual would be correlated so individuals should only be included once. We did not enter the expert into the trial for a second time. The protocol was

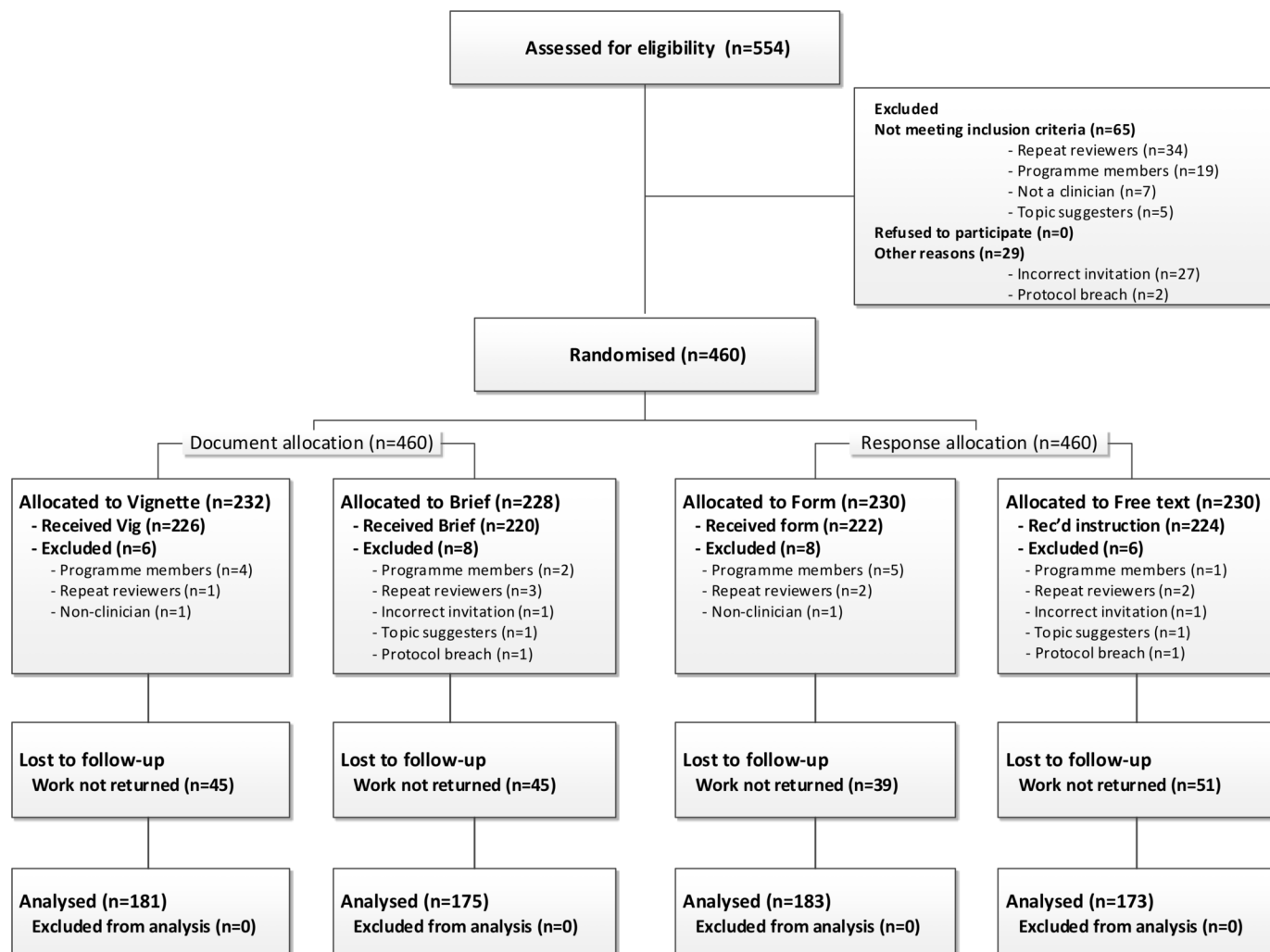


Figure 1 Consort diagram of participant flow.

updated to make it clear that only the data relating to the first vignette that a trial participant commented on during the study would be used.

We also developed a procedure to respond to reviewer queries in a standardised way, to ensure that participants received correct information about the review process within the trial. The procedure was worded in such a way that reviewers remained unaware of the trial hypothesis.

RESULTS

The flow of participants through the study is shown in [figure 1](#). Of the 460 randomised participants, 232 were allocated to receive the vignette and 228 were to receive the commissioning brief; 230 were allocated to a structured response and 230 were to free text.

A total of 356 participants provided a response within the time required to affect the decision of the programme, and 90 did not. Fourteen participants were identified after randomisation as not eligible and were excluded from the trial at allocation stage.

Internal feasibility

We were able to randomise participants, and the quality scores of the first third of reviewer comments were above the stopping threshold. The study therefore continued to recruit for the planned year.

Primary outcome

The distribution of scores assigned by the assessors is shown in [figure 2](#).

Counting non-responders as scoring 0, the pooled mean quality score was 2.4, with an SD of 0.95.

The commissioning brief scored 0.037 (Cohen's $d=0.039$) extra quality points over the vignette arm; the structured form response scored 0.335 (Cohen's $d=0.353$) over the free text. There were no interactions between the allocations ($p=0.730$).

As a sensitivity analysis, we repeated this process, omitting non-responders. The pooled mean quality score without the non-responders was 3.0, with an SD of 0.81. Using data from only responders, the commissioning brief scored 0.06 (Cohen's $d=0.071$) quality points over the vignette; the structured response scored 0.25

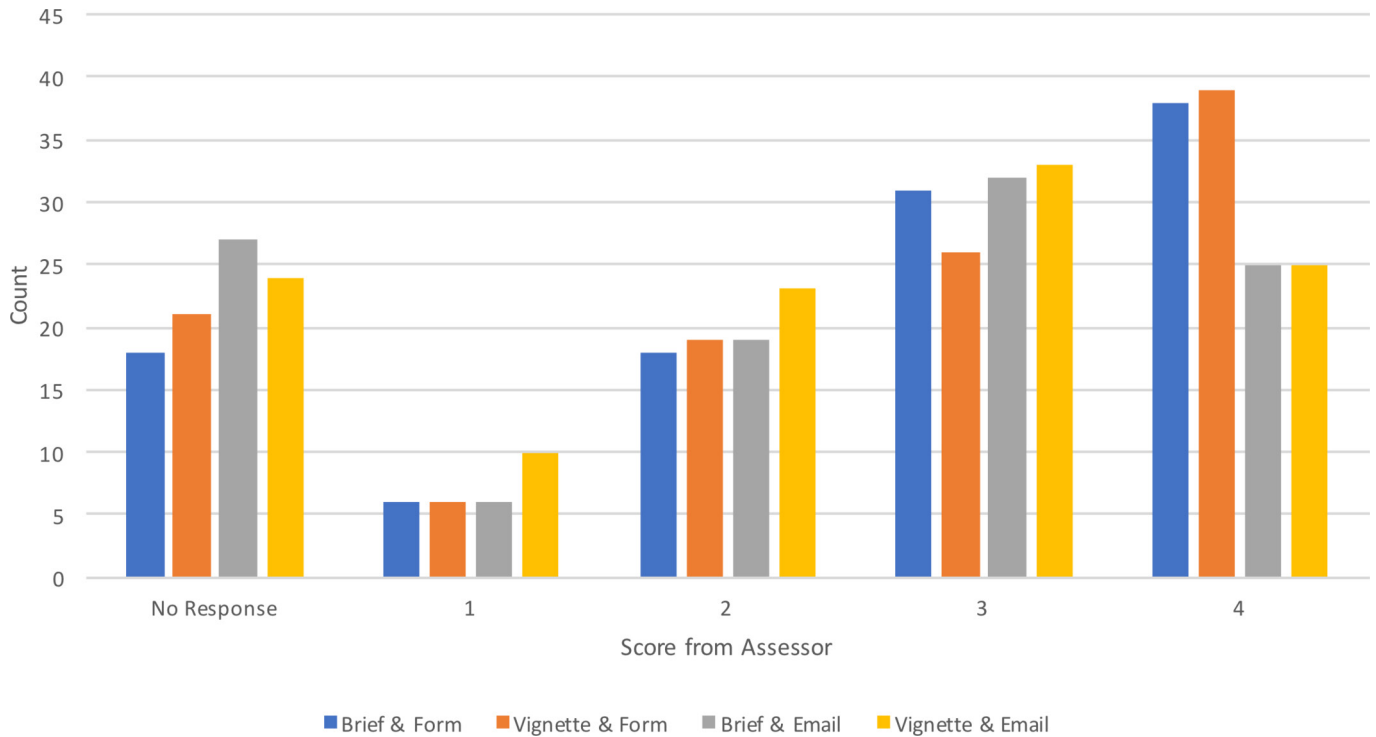


Figure 2 Assessor scores across allocated groups.

(Cohen's $d=0.309$) over a free text response. There were no interactions between the allocations ($p=0.524$). The effect was smaller but still over the predefined threshold for a worthwhile effect.

There was therefore no important difference between the allocation to receive either the commissioning brief or the vignette, but a response using a structured form appears to show a worthwhile (using the predefined criterion) benefit over a free text response.

Quality of allocation concealment — vignette versus commissioning brief

Table 1 sets out the analysis of masking. It appears that the assessors were not completely masked, but the excess correct guess rate was small. As the assessors were better able to identify allocation when just the commissioning brief was sent, it seems that this is driven by a failure to comment on items included in the vignette but not in the commissioning brief.

Effect of randomised allocation on likelihood of response

We explored whether any of the allocations had an impact on the willingness of an expert to complete the requested work. This is important as if any of the allocations were

actively off-putting then a lack of willingness of experts to participate might offset any benefit of higher quality responses from those who did return opinions.

Using the allocation figures and the analysed figures from figure 1, a four-sample test for equality of proportions gives a p value of 0.72. We therefore conclude that there is no relationship between allocation of either material or response and the likelihood that an expert returns their comments.

Willingness to participate in the review process

To address this question, we drew on routine data used within the HTA programme. In 2014, clinical experts were approached on 1338 occasions to contribute to vignettes. On 555 occasions, there was no response to the request. On 281, the opportunity was declined. The remaining 502 resulted in an acceptance of the invitation. This is a larger figure than the 460 randomised experts, as 42 were approached to review two or more different vignettes during the course of the study, and only the first acceptance was included in the randomised trial.

We prospectively identified six groups of sources from which these experts had been identified.

Table 1 Assessment of masking to vignette or commissioning brief allocation

Subset	Correct guess of allocation	Incorrect guess of allocation	Significance*
All responses	222 (62.4%)	134 (37.6%)	$p<0.00001$
Received vignette	104 (57.4%)	77 (42.6%)	$p=0.053$
Received commissioning brief	118 (67.4%)	57 (32.6%)	$p<0.00001$

*Binomial test, assuming that, if masking were perfect, the correct guess rate would be 0.5, see Statistical methods section.

Table 2 Sources of experts

	Accepted	Declined	No response	Total
NETSCC internal databases	363	217	303	883
External databases	56	24	183	263
Recommendation	56	24	37	117
Search engines	2	1	5	8
Other source	9	3	8	20
Unknown	16	12	17	45
Total	502	281	553	1336

‘NETSCC internal databases’ refers to records which NETSCC keeps of people who have previously worked with NIHR programmes. ‘External databases’ includes sources such as Specialist Info (<http://specialistinfo.com>) which keep records of clinical expertise. ‘Recommendations’ occur when a particular expert is suggested to the programme to review a vignette, usually by a TIDE panel member. ‘Search engines’ refers to generic internet search engines such as Google and Duck Duck Go. ‘Other source’ includes a mixture of small volume sources such as NICE committees. Occasionally, we have no record of the source from which an expert was identified, and these are classified as ‘Unknown’.

While not in the original analysis plan, we have explored the relationship between the likelihood that an expert works with the programme to the source from which they were identified.

A χ^2 test across the whole table has a p value of less than 0.001, implying a relationship between the source of an expert and them completing a review. We investigated further by amalgamating pairs of columns. Testing responders (people who did the work and people who positively declined) against non-responders gives a p value of less than 0.001. Conversely, testing people who did the work against those who did not (decliners and non-responders) gives a non-significant p value of 0.076.

Table 2 contains data from all occasions when a clinician was invited to review. That means some clinicians are included more than once. It is common when finding reviewers for this programme that clinicians decline because of workload but accept when invited for a further vignette. We therefore considered it reasonable to include all invitations in this table. As a sensitivity analysis, we repeated the χ^2 tests removing duplicate invitations, thus reducing the total count of the ‘Accepted’ column to 460. There was no change in the p values when expressed to two significant figures.

It is clear from **table 2** that experts who are already known to NETSCC are far more likely to respond to a request for help than those who are not. Experts who are recommended by their peers are also more likely to respond positively. The ‘other’ category also had a high response rate, but the absolute numbers here are small so we are reluctant to draw a conclusion. When the invitation is responded to, there is no significant difference in

the likelihood that the expert will complete the offered task. We therefore conclude that experts drawn from sources where we would expect them to be familiar with the programme are more likely to contribute than those who are less likely to know of this funder.

Post hoc analysis — primary outcome

One of the journal referees suggested that it may be more appropriate to consider the primary outcome measure as ordinal data rather than ratio, due to the narrow range of the scale. We considered this in a post hoc analysis. All the allocations had a median quality score of 3, with an IQR of 2–4.

The appropriate test of significance then becomes the Mann-Whitney U test. The results of the significance test are shown in **table 3**, for both our preferred approach of scoring non-responders as 0 and for excluding non-responders. Significance is maintained in the mode of response, and it is still not present in the document allocation.

We have assessed the effect size in this model using rank-biserial correlation.¹⁴ We have not considered the effect size in the document allocation as there was no significant difference. The effect size in the response allocation was 0.140 when no response scored a 0 and was 0.138 where non-responders are ignored. These correlations would usually be viewed as very small.

DISCUSSION

NETSCC has had a research on research programme for several years, undertaking research to improve delivery of NIHR programmes, to document their influence and to reduce waste.^{15–21} This is however the first randomised trial of the research funding process to take place within NIHR. As such, it served two purposes — first to investigate

Table 3 Significance tests for the effects of document and response allocation, using the Mann-Whitney U test

Allocation	Non-responders score 0 included	Non-responders not included
Document	p=0.767	p=0.568
Response	p=0.008	p=0.018

the question around how best to involve clinical experts and second to demonstrate that a randomised trial is possible inside this research funding organisation.

There is a significant literature on the use of reviewers for the evaluation of journal articles, a few publications on using reviewers to assess funding applications, but nothing on the best way to involve clinical experts in a commissioned mode funding programme.

We have shown in this study that the material sent to reviewers to assess appears to have no consequence on the usefulness of the comments which reviewers provide, but the format in which they are asked to provide those comments is important. However, this conclusion needs to be viewed with caution.

While the assessors were reasonably masked to allocation with regard to the material distributed, it was implausible to mask them to the means of response within the resources available. This means that the comparison where we have shown a meaningful difference was unmasked—and the assessors preferred the condition which most matched current practice. When we reanalysed the data using a non-parametric model, the level of correlation between response allocation and quality score was small—lower than would usually be viewed as meaningful. This may indicate that using a structured form is superior or just that the assessors were used to evaluating and using responses received this way and so they rated these responses higher. The assessors (HTA staff) reviewing the material received considered that there may also be an element of professional group characteristics in the usefulness of comments provided via different formats. That is, certain professional groups tend to provide longer comments than others and this was more pronounced in the free text form, which made some of the reviews difficult to handle and to interpret. This was drawn from experience, rather than information available within the trial.

Conversely, for the adequately masked comparison, no difference was shown in the primary outcome. We found this surprising. The investigators' prior hypothesis (unlike that of the TIDE panel chairs who suggested this question) was that providing more information would lead to a more useful response from the reviewers.

It is reassuring that the material and response allocations appear to have no effect on an expert's willingness to provide their opinion, if experts actively did not engage with any of the options that would rule them out in practice.

There is a need to further investigate how assessors are reviewing the material provided by reviewers and how reviewers interact with the material provided. We are currently planning this qualitative work.

The work exploring the willingness of experts sourced through various routes provided the unsurprising conclusion that experts who are familiar with the programme are more likely to respond than experts with little exposure to NIHR and the HTA programme. In a world where clinicians are often continually bombarded with requests to contribute to various activities which they do not view as part of their core job, this was to be expected. It may have implications for NIHR's communications

strategy—highlighting that the awareness of NIHR in the clinical community in the UK may result in more clinicians willing to review research ideas.

This trial highlighted the need for a research process for future studies set within this research funder. This work was completed by interested people in their 'spare time'. This has had consequences both for the timeliness of reporting and for the work which we were able to undertake. Ideally, a process evaluation to explore how assessors and reviewers interact with the materials provided would have taken place in parallel to the quantitative trial—but this was not possible within the resources available. This study has unearthed questions of interest to the organisation, although no resource has been found as of yet to follow-up on these questions.

The approaches used here could be reproduced to look at other uses of clinical reviewing. This would be relevant to NETSCC and also potentially relevant to other funders—all of which use reviewing to help assess grant proposals, but few if any have a similar process for prioritising research questions.

Twitter Andrew Cook: @ajcook

Acknowledgements We gratefully acknowledge constructive advice from our trial steering group (Paula Barratt, Louise Craig, Peter Davidson, Tom Kenny, Sarah Puddicombe and Karen Williams). We thank the NETSCC prioritisation team for facilitating this work and the panel researchers and consultant advisers who assessed the material received from the expert reviewers. We thank Yoon Loke and Clare Wilkinson who made the original observation which inspired this study. We would also like to thank the University of Southampton Faculty of Medicine Ethics Committee for their advice, the clinical experts who unknowingly contributed to this work and the referees appointed by the Journal.

Contributors AC designed the study, analysed and interpreted the data, drafted the article and approved the final version for publication. ES designed the study, collected the study data, interpreted the data, critically reviewed the article and approved the final version for publication. GD collected the study data, critically reviewed the article and approved the final version for publication.

Disclaimer The views and opinions expressed herein are those of the authors.

Competing interests All authors are employed by the University of Southampton to contribute to the National Institute for Health Research (NIHR). Their continuing employment may to some extent depend on the continued funding of NIHR.

Ethics approval University of Southampton Faculty of Medicine Ethics Committee #8192.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Anonymised data may be requested from the corresponding author.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;374:86–9.
2. Walley T. Health technology assessment in England: assessment and appraisal. *Med J Aust* 2007;187:283–5.

3. Raftery J, Powell J. Health Technology Assessment in the UK. *The Lancet* 2013;382:1278–85.
4. Health Technology Assessment (HTA) Programme Website. <http://www.nets.nihr.ac.uk/programmes/hta> (accessed 17 May 2016).
5. Smith R. Classical peer review: an empty gun. *Breast Cancer Res* 2010;12(Suppl 4):S13.
6. Marsh HW, Jayasinghe UW, Bond NW. Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. *Am Psychol* 2008;63:160–8.
7. Wessely S. Peer review of grant applications: what do we know? *Lancet* 1998;352:301–5.
8. Herbert DL, Graves N, Clarke P, et al. Using simplified peer review processes to fund research: a prospective study. *BMJ Open* 2015;5:e008380.
9. Moher D, Simera I, Schulz KF, et al. Helping editors, peer reviewers and authors improve the clarity, completeness and transparency of reporting health research. *BMC Med* 2008;6:13.
10. Hodgson C. How reliable is peer review? An examination of operating grant proposals simultaneously submitted to two similar peer review systems. *J Clin Epidemiol* 1997;50:1189–95.
11. Hoffmann TC, Bennett S, Tomsett C, et al. Brief training of student clinicians in shared decision making: a single-blind randomized controlled trial. *J Gen Intern Med* 2014;29:844–9.
12. Cook J, Hislop J, Adewuyi T, et al. Assessing methods to specify the target difference for a randomised controlled trial: DELTA (Difference ELicitation in TriAls) review. *Health Technol Assess* 2014;18:1–175.
13. R Core Team. R: a language and environment for statistical computing.
14. Cureton EE. Rank-biserial correlation. *Psychometrika* 1956;21:287–90.
15. Bhurke S, Cook A, Tallant A, et al. Using systematic reviews to inform NIHR HTA trial planning and design: a retrospective cohort. *BMC Med Res Methodol* 2015;15:1.
16. Turner S, Bhurke S, Cook A. Impact of NIHR HTA Programme funded research on NICE clinical guidelines: a retrospective cohort. *Health Res Policy Syst* 2015;13:37.
17. Turner S, Wright D, Maeso R, et al. Publication rate for funded studies from a major UK health research funder: a cohort study. *BMJ Open* 2013;3:e002521.
18. Turner S, Wright D, Maeso R, et al. Avoiding waste of research resource: cohort study of publication rate for funded studies from a major UK research funder. *Trials* 2013;14:P95.
19. Raftery J, Young A, Stanton L, et al. Clinical trial metadata: defining and extracting metadata on the design, conduct, results and costs of 125 randomised clinical trials funded by the National Institute for Health Research Health Technology Assessment programme. *Health Technol Assess* 2015;19:1–138.
20. Chinnery F, Young A, Goodman J, et al. Time to publication for NIHR HTA programme-funded research: a cohort study. *BMJ Open* 2013;3:e004121.
21. Douet L, Milne R, Anstee S, et al. The completeness of intervention descriptions in published National Institute of Health Research HTA-funded trials: a cross-sectional study. *BMJ Open* 2014;4:e003713.