BMJ Open

Involving Clinical Experts in Prioritising Topics for Health Technology Assessment: A Randomised Controlled Trial

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016104
Article Type:	Research
Date Submitted by the Author:	27-Jan-2017
Complete List of Authors:	Cook, Andrew; University of Southampton, Wessex Institute; University Hospital Southampton Streit, Elke; University of Southampton, NIHR Evaluation, Trials and Studies Coordinating Centre Davage, Gill; University of Southampton, Wessex Institute
Primary Subject Heading:	Medical publishing and peer review
Secondary Subject Heading:	Evidence based practice, Health services research
Keywords:	peer review, Health Technology Assessment, clinicians

Involving Clinical Experts in Prioritising Topics for Health Technology Assessment: A Randomised Controlled Trial

Andrew Cook [1,2 *], Elke Streit [3], Gill Davage [3]

1 Wessex Institute, University of Southampton, UK
2 Southampton University Hospitals Foundation Trust, UK
3 NIHR Evaluation Trials and Studies Coordinating Centre, University of Southampton, UK

* Corresponding Author

Clinical Trials Registration: ANZCTR12614000167662 (Retrospectively Registered)
Ethics Approval: University of Southampton, Faculty of Medicine Ethics Committee 8192

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 2 types of review materials (long document vs. short document) and response modes (structured review form vs. free-text email response). Previous and current members of the funder's programme groups were excluded. Response quality was assessed by use of a 4-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 3.0 (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 0.335 (Cohen's d = 0.353) over free text. Experts allocated to structured response were more likely to provide comments than those allocated to free-text (effect size 0.36 standard deviations). There were no interactions between the allocations (p = 0.730).

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: ANZCTR12614000167662

Article Summary

Strengths and limitations of this study

- The trial included all eligible clinical experts over the course of a year
- The strongest effects were shown in areas where assessors could not be masked
- The findings will directly influence practice in a major clinical trials funder

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 the right research questions are asked; 2) ensuring that study designs are appropriate and are of methodological quality; 3) ensuring the findings from funded research are available in the public domain; 4) ensuring that funded research is unbiased and usable [1].

The NIHR Health Technology Appraisal (HTA) programme was established in the 1990s, in part to address market failure in UK health research, and is now imbedded in the National Institute for Health Research (NIHR), managed by the NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC). The programme is the major public funder of pragmatic trials in the United Kingdom, and it's 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 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 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 (or TIDE) panel. These are standing groups of up to twenty clinicians and lay members, grouped by 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 - 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 underway) 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 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 on to the TIDE panel for consideration. Sometimes the secretariat will iterate a point with the clinical expert.

Around thirty percent 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 had to consider what evidence we needed in order to refine the processes which we use to develop the research questions 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 five to ten minutes, and rather than asking respondents to complete a form, the programme would accept responses as a reply to the initial invitation email. It was thought 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 or a commissioning brief. The other between being asked to respond using free text, or 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 paper [11] reporting a trial evaluating training for medical students, and 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 two months into our one year recruitment period.

Participants and sample size

The participants were every clinical expert 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 554 clinical experts were approached to comment on possible research and of these 460 were randomised.

For experts approached to contribute to more than one topic during the recruitment period, only their involvement with the first topic was included in the study. Experts were also excluded if they were current or previous members of HTA programme groups – such as the TIDE panels or funding boards. Contributors were also excluded 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 size 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 underway, but 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 governing ethics committee.

HTA staff assessing the responses received were aware of the hypotheses being tested, but were not informed of the allocation of participants who provided the responses that they were assessing. However, whether the response was provided as 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 to 4) 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 the expert and likelihood of responding

We planned to assess whether assessors could identify whether the participants had been sent a vignette or a commissioning brief, to assess the quality of masking.

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 ANOVA, with the quality being the response variable, and the two allocations (vignette vs. commissioning brief, and free text vs. form) as the input variables. Interaction was investigated.

The likelihood of response was modelled using the binomial distribution. For assessment of masking, p values were calculated using a binomial test, assuming that if masking were perfect the correct guess rate would be 0.5.

Sources of data

Data on vignette allocation and quality of responses received was collected specifically for this study. Data on expert's willingness to participate in the reviewing process was 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 to 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 broad, and a one point difference in the scale was not well understood. We revised the scale to 0-4 and rescored the initial set of responses, and out reviewers found this much more satisfactory. Under both systems reviewers were allowed to express fractional values.

We modified the inclusion criteria several times during the course of the study, usually to make them more restrictive. This was always precipitated by a situation arising which was both (i) unanticipated and (ii) likely to undermine the study assumption that randomised individuals would behave independently.

We had to refine our definition of a clinical expert (as opposed to a methodological expert), and 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

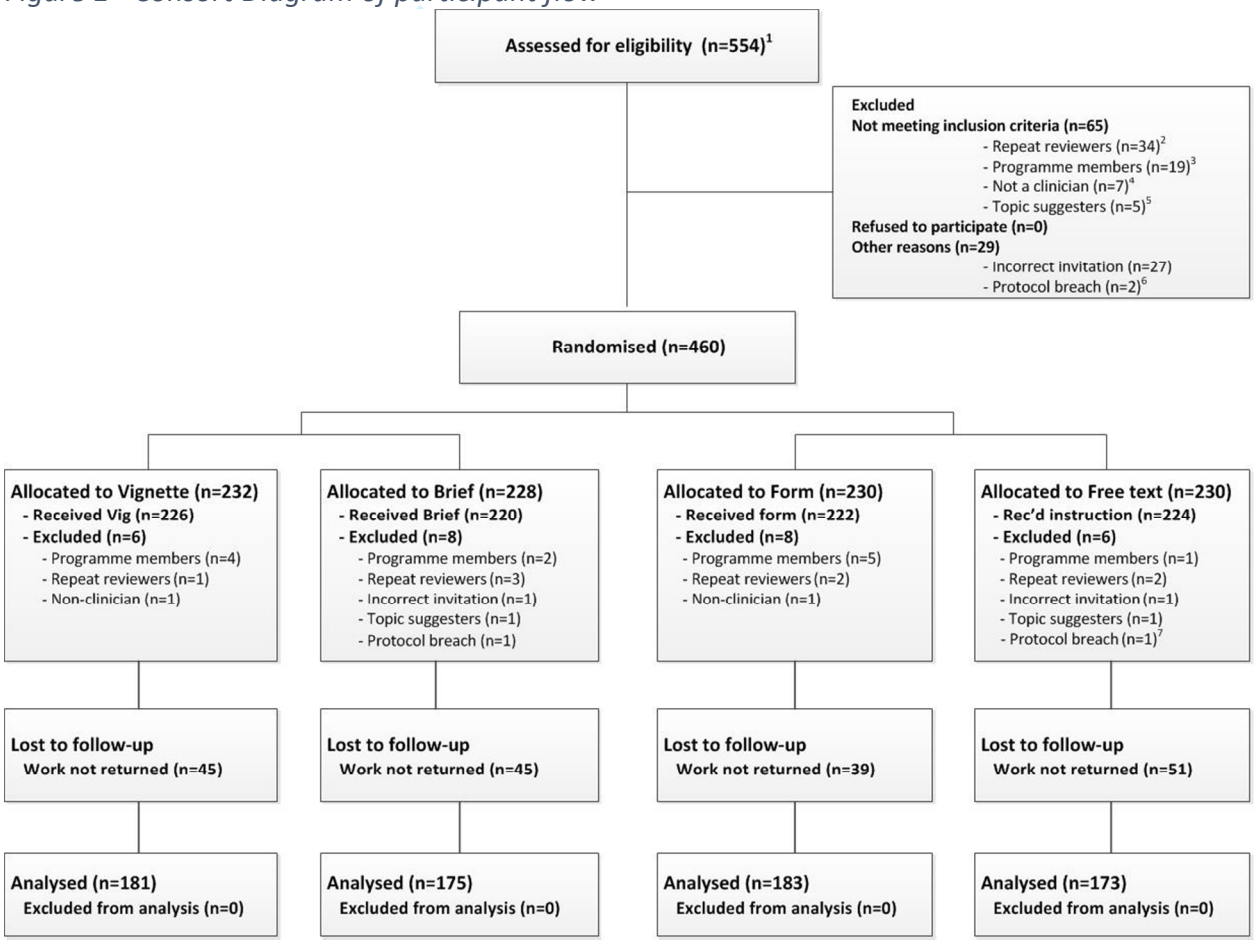
queries in a standardised way, to ensure participants received correct information about the review process within the trial. The procedure was worded in such a way that reviewers remained unaware of the trial hypothesis.

Results

The flow of participants through the study is shown in *Figure 1*. Of the 460 participants, 232 were allocated to receive the vignette and 228 the commissioning brief; 230 were allocated to a structured response and 230 to free text.

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.

Figure 1 - Consort Diagram of participant flow



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 pooled mean quality score was 3.0, with a standard deviation of 0.95.

Counting non-responders as scoring 0: the commissioning brief scored 0.037 (Cohen's d = 0.039) extra quality points over the vignette arm; and the structured form response 0.335 (Cohen's d = 0.353) over the free text. There were no interactions between the allocations (p = 0.730).

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 benefit over a free text response.

Quality of Allocation Concealment – vignette vs. commissioning brief

Table 1 - Assessment of Masking to Vignette or Commissioning Brief Allocation

Subset	Correct guess of allocation	Incorrect guess of allocation	Correct Guess Rate	Significance
All Responses	222	134	0.624	p < 0.00001
Received Vignette	104	77	0.574	p = 0.053
Received Commissioning Brief	118	57	0.674	p < 0.00001

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.

Willingness to participate in the review process

To address this question we drew on routine data used within the HTA programme. In 2014, 1338 clinical experts were approached to contribute to vignettes. 555 did not respond to the request at all. 281 responded, declining the opportunity. The remaining 502 accepted the invitation to contribute. This is a larger figure than the 460 randomised experts, as 42 were approached more than once to review during the course of the study, and only the first acceptance was included in the randomised trial.

We prospectively identified 6 groups of sources from which these experts had been identified

Table 2 Sources of experts

	Accepted	Declined	No Response	Total	Acceptance Rate	SD Acceptance Rate
NETSCC Internal	363	217	303	883	*41.1%*	1.7%

Databases						
External Databases	56	24	183	263	*21.3%*	2.5%
Recommendation	56	24	37	117	*47.8%*	4.6%
Search Engines	2	1	5	8	*25.0%*	15.3%
Other Source	9	3	8	20	*45.0%*	11.1%
Unknown	16	12	17	45	*35.6%*	7.1%
Total	**502**	**281**	**553**	**1336**	***37.6%***	**1.3%**

'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'.

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.

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 [13-19]. This is however the first randomised trial of the research funding process to take place within NIHR. As such it served two purposes – firstly to investigate the question around how best to involve clinical experts; but also 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 regards 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. 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 rated these responses higher. The researchers 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

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 providing more information would lead to a superior response from the reviewers.

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.

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 has been able to be undertaken. 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.

Acknowledgements

We gratefully acknowledge constructive advice from our trial steering group (Paula Barratt, Louise Craig, Peter Davidson, Tom Kenny, Sarah Puddicombe, Karen Williams). We thank the NETSCC prioritisation team for facilitating this work. We would also like to thank the Southampton University Faculty of Medicine Ethics committee for their advice, and the clinical experts who unknowingly contributed to this work.

Author's Contributions

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.

References

1 Chalmers I, Glasziou PP. Avoidable waste in the production and reporting of research evidence. 2009;**374**:86–9. doi:10.1016/S0140-6736(09)60329-9

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. doi:10.1016/S0140-6736(13)61724-9

4 Health Technology Assessment (HTA) Programme Website. nets.nihr.ac.uk. http://www.nets.nihr.ac.uk/programmes/hta (accessed 17 May2016).

5 Smith R. Classical peer review: an empty gun. *Breast Cancer Res* 2010.

6 Marsh HW, Jayasinghe UW, Bond NW. Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist* 2008;**63**:160–8. doi:10.1037/0003-066X.63.3.160

7 Wessely S. Peer review of grant applications: what do we know? *The Lancet* 1998;**352**:301–5. doi:10.1016/S0140-6736(97)11129-1

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. doi:10.1136/bmjopen-2015-008380

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. doi:10.1186/1741-7015-6-13

10 Hodgson C. How reliable is peer review? An examination of operating grant proposals simultaneously submitted to two similar peer review systems. *Journal of clinical epidemiology* 1997.

11 Hoffmann T, 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. doi:10.1007/s11606-014-2765-5

12 Cook JA, Hislop J, Adewuyi TE, *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**:v–vi–1–175. doi:10.3310/hta18280

13 Bhurke S, Cook A, Tallant A, *et al.* Using systematic reviews to inform NIHR HTA trial planning and design: a retrospective cohort. *BMC medical research methodology* 2015;**15**:1. doi:10.1186/s12874-015-0102-2

14 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. doi:10.1186/s12961-015-0025-8

15 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. doi:10.1136/bmjopen-2012-002521

16 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* Published Online First: November 2013. doi:10.1186/1745-6215-14-S1-P95

17 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. doi:10.3310/hta19110

18 Chinnery F, Young A, Goodman J, *et al.* Time to publication for NIHR HTA programme-funded research: a cohort study. *BMJ Open* 2013;**3**:e004121–1. doi:10.1136/bmjopen-2013-004121

19 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. doi:10.1136/bmjopen-2013-003713

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors

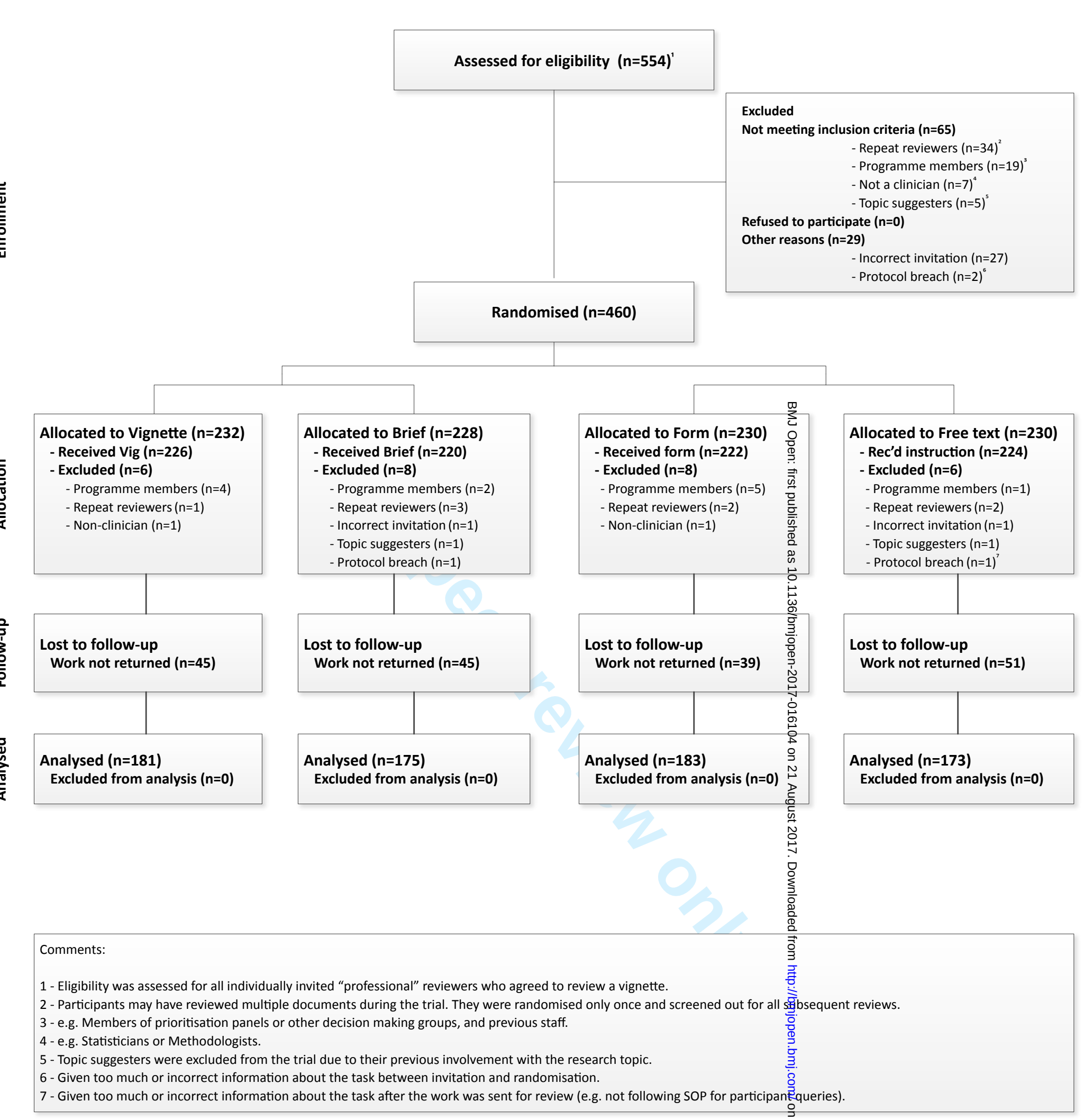
Disclaimer: The views and opinions expressed herein are those of the authors

Competing Interests: All authors are employed by the Universty of Southamption, to contribute to NIHR. Their continuing employment may to some extent depend on the continued funding of NIHR

Participant Consent: Not obtained, with the agreement of the University of Southampton Faculty of Medicine Ethics Committee (#8192)

Data Sharing Statement: Anonymised data may be requested from the corresponding author

Participant flow diagram



Comments:

1 - Eligibility was assessed for all individually invited "professional" reviewers who agreed to review a vignette.
2 - Participants may have reviewed multiple documents during the trial. They were randomised only once and screened out for all subsequent reviews.
3 - e.g. Members of prioritisation panels or other decision making groups, and previous staff.
4 - e.g. Statisticians or Methodologists.
5 - Topic suggesters were excluded from the trial due to their previous involvement with the research topic.
6 - Given too much or incorrect information about the task between invitation and randomisation.
7 - Given too much or incorrect information about the task after the work was sent for review (e.g. not following SOP for participant queries).

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	4-6
	2b	Specific objectives or hypotheses	6
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	7
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	10
Participants	4a	Eligibility criteria for participants	7-8
	4b	Settings and locations where the data were collected	7-8
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	5-6, 7
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	9
	6b	Any changes to trial outcomes after the trial commenced, with reasons	10
Sample size	7a	How sample size was determined	7-8
	7b	When applicable, explanation of any interim analyses and stopping guidelines	10
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	8
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	8
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	8
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	8
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	8

		assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	n/a
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	9
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	n/a
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	12
	13b	For each group, losses and exclusions after randomisation, together with reasons	12
Recruitment	14a	Dates defining the periods of recruitment and follow-up	7-8
	14b	Why the trial ended or was stopped	7-8
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	n/a
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	12, 9
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	12-14
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	n/a
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	12-14
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	n/a
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	16
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	16-17
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	15-17
Other information			
Registration	23	Registration number and name of trial registry	1
Protocol	24	Where the full trial protocol can be accessed, if available	By email
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	20

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

Project Protocol

What is the optimum way to approach external clinical vignette reviewers for the NIHR Health Technology Assessment programme: a factorial RCT

12th October 2013
Version 0.4

Change Control

Date	Description of Change	Made By
30th May 2013	First Draft	Andrew Cook
9th July 2013	Following first steering group meeting • factorial design • historical cohort to address invitation wording • Can 'use' assessors be blinded?	Andrew Cook
5th October 2013	Updates to the analysis plan following internal discussion	Andrew Cook
7st October 2013	Updates for external consumption	Andrew Cook
9th October 2013	Changes to expert approach process following first ethics review	Andrew Cook

Abstract

The HTA programme consumes large numbers of clinical experts to inform its commissioning briefs. This takes large amounts of staff time internally, and that of the clinical experts externally.

At a recent induction meeting, two new Panel Chairs suggested that clinical experts do not need an entire vignette in order to make useful and constructive comments - and may be put off by invitation letters which suggest an hour of their time might be needed. He suggested that the 'commissioning brief' portion may be sufficient - and could be processed by the expert in much less time. In addition, if we could offer work which takes less time, we may get a higher uptake rate.

If this is true, it would help optimise the HTA programme's use of clinical experts. We therefore propose a trial of approaches to clinical experts for vignettes using a factorial design to assess two questions

- What is the effect of sending the 'commissioning brief' rather than the vignette to experts for assessment?
- What is the effect of asking experts to respond with a free text response in email, compared to completing a structured form?

Introduction

The NIHR HTA programme commissions research through two main routes:

1. A researcher led route, where investigators propose research questions and methods for answering them to the programme, which then uses its processes to decide which proposals to fund
2. A commissioned route, where the programme identifies research questions of importance to the NHS through a number of processes (literature review, consultations with stakeholders etc) and then asks panels of clinical and lay experts to validate prioritisation decisions. These panels are provided with briefing documents of somewhere between 4 and 10 pages (known internally as vignettes) to help them take these decisions. For those topics which are prioritised a shorter document (known as a commissioning breif) is advertised to the research community. As part of the preparation of these documents external clinical experts are involved to provide information on clinical practice and clinical unknowns.

The HTA programme consumes large numbers of clinical experts to inform its commissioning briefs. This takes large amounts of staff time internally, and that of the clinical experts externally.

Current practice

- Experts are identified
- Experts are approached with offers of work to review the vignette, with an expectation that the task will take around an hour
- The vignette and assessment form are delivered to the expert as email attachments
- Experts are requested to complete the PDF form and return it to the office.

This process represents a fair amount of friction - consumes an hour, working with PDF and attachments - and it is unsurprising that a significant fraction of experts decline to engage.

At a recent induction meeting, two new panel chairs suggested that clinical experts do not need an entire vignette in order to make useful and constructive comments - and may be put off by invitation letters which suggest an hour of their time might be needed, and excessive attachments. He suggested that the 'commissioning brief' portion may be sufficient - and could be processed by the expert in much less time. In addition, if we could offer work which takes less time, we may get a higher uptake rate.

This suggests an alternative process where

- Experts are identified

- Experts are approached with offers of work to review the draft commissioning brief, with an expectation that the task will take around 10 minutes.
- The brief is delivered as the body of an email, with any supplemental questions which the CA or PR wish to ask also in the body of the email.
- The expert responds by replying to the email
- We would offer the vignette if people want it.

This process removes friction, and gives the expert less work, so we might expect more experts to engage with this process. Increasing clinical participation in refereeing would help the programme to demonstrate engagement with the clinical community, but it would only be of use if quality is maintained.

Importance

Expert identification takes up a significant internal resource. If we were able to improve the response rate of experts without decreasing the quality of their input then we may be able to free resource in expert identification. We would also reduce the overall burden on our pool of clinical experts.

Governance

Steering Group

Confirmed Members

- Andrew Cook (CI) - CPHM & Fellow in HTA
- Elke Streit - Panel Researcher
- Peter Davidson - Director of HTA, NETSCC
- Paula Barratt - Senior Research Fellow
- Tom Kenny - Director of External Relations, NETSCC
- Louise Craig - Panel Manager (until December 2013)
- Karen Williams (From December 2013)

Ethics

Will need approval from University of Southampton Faculty of Medicine Ethics Committee

Registration

Will ideally be registered, but no registry yet found which will take an unfunded trial involving neither patients nor a health outcome.

Research Questions

1. What is effect on completion rate and usefulness of responses received when clinical experts in the vignette process are provided with the 'commissioning brief' rather than the vignette for comment.
2. What is the effect on completion rate and usefulness of responses when clinical experts in the vignette process are asked to respond using free text, rather than completing and returning a structured form.

Definitions

Acceptance Rate	The proportion of requests where the work is accepted.
Completion Rate	proportion of experts approached who complete and return the requested work within agreed timescales
Commissioning Brief	The front portion of the vignette, up to and including the 'background box', which will later in the process be used as the basis for the commissioning brief to be advertised.
Usefulness of responses	A decision as to which set of responses is more useful, as assessed jointly by the CA & PR responsible for producing a vignette.

PICOs

Factorial Trial

P	Invited Clinical Experts
I	**1) Commissioning Brief** **2) Email Response**
C	1) Vignette 2) Form
O	a. Return an opinion within time limit b. Usefulness c. Requests for vignette in I1 box **d. Time to respond**

Theoretical Framework

The questions will be investigated using a 2x2 factorial randomised controlled trial Individuals approached to participate will not be informed that other participation options are available (other than those sent the 'brief' will be offered the full vignette on request - they will not however be told there is an alternate arm).

We're interested in assessing how well experts engage with the programme. Hiding the existence of other arms from the experts will prevent them favouring any part of the documentation they receive (such as the 'brief' component) or trying to anticipate what other modes of response might have been offered to them. We believe this is reasonable is the default position for experts is to decline to engage - we will here be testing how different approaches affect that engagement rate - and having those approached know there is an assessment process going on will confound the results.

We do not believe that results for the two allocations - alternate documents, and alternate means of response - will interact, but we will test this as part of the assessment.

Pilot Study

We have conducted a pilot study internal to NETSCC (ie not affecting engagement with people outside NETSCC) to assess

- The record keeping system for randomisation and results
- The practicalities of randomisation

Randomisation

Randomisation will be by precalculated blocks.

We have demonstrated that the blocks can be demonstrated, and that experts can be allocated to a randomisation by an assistant programme manager using a set of work instructions.

We have not yet acted on this randomisation to send any material to experts.

Record Keeping

The main record keeping document for the factorial trial will be a table, with experts recorded on rows, with columns for

- Expert name or ID code
- Document allocation
- Response allocation
- Date of work allocation
- Date of work received (if any)
- Whether the expert requested the vignette

We have demonstrated that

1. An assistant programme manager can enter experts into the table in the order they accept work can be done, in accordance with a set of work instructions
2. The table can be pre-populated with allocation codes
3. The relevant dates can be entered into the table by the appropriate programme manager

Inclusion and Exclusion Criteria

Vignettes

Inclusion

- All vignettes prepared during 2014 rounds.

Except

- Vignettes which ask a methodological, rather than clinical, question.
- Vignettes where the responsible CA does not want external experts.
- Vignettes which have been through previous external review, other than as a panel topic.

Experts

Inclusion

- Clinical experts approached to advise on included vignettes

Except

- Non-clinical experts. Clinical-Academics count as clinical experts. The final decision as to whether an advisor is 'clinical' rests with the Steering Group responsible for the vignette, who may take advice from the relevant CA.
- People who have provided advice on the panel topic from which the vignette was derived
- Current and previous members of HTA prioritisation and TIDE panels

Sampling and Randomisation

Experts will be sampled in line with the usual practice at NETSCC.

Experts who are approached to be offered work will be sent an invitation letter including the text NETSCC is currently assessing what material should be sent to clinical reviewers, therefore please do not be surprised if the material you are sent, or the accompanying instructions, differ from those you may have received in the past.

Within the randomised portion of the study, allocations to each of the two randomisations (document and response) will be pre-generated. Randomisation will be by blocks, with random block sizes of 2, 4 and 6.

Experts will be allocated based on the order their acceptances of work are received, as defined by the time stamp of the acceptance email.

Data Collection

Acceptance and Completion

For the vignette rounds of 2014 we will prospectively collect for each expert approached

1. Panel
2. Vignette ID
3. Vignette Round
4. Was the work accepted
5. Allocated to Vignette or Brief document
6. Allocated to Form or Email response
7. Was the work completed
8. Date work sent
9. Date response received

How useful are the responses?

Each CA & PR team scores each response received on a scale of 0-10. Fractional scores are allowed (eg if a CA score 7 and a PR score 8, the summary result could be 7.5)

Also ask why the grade was given.

Pick list:

- Response ignored
- Told us stuff we knew already
  - From literature
  - From another expert
- Changed the story of the vignette
- Changed the facts
- Changed the science
- Changed the PICO
- Other (specify)

Stopping Rules

Safety

After each vignette round, CA/PR pairs will be asked for each topic

"Was the overall information received from experts to inform your decisions at least as good as you would normally expect".

With a response of 'no' for more than 25% of topics, the steering group will be asked to consider whether the quality of commissioning briefs is at risk, and if so whether the trial should be stopped.

Analysis Plan

Primary Outcome

The main outcome of interest is the usefulness of the response received. This is measured on a scale of 0 to 10, with no response scoring 0. Fractional values are allowed.

The mean and standard deviations of the usefulness score will be calculated

Will be modelled with ANOVA, with usefulness of response being the response variable, and the two trial allocations being the input variables.

We will investigate interaction between the allocations.

Secondary Outcomes

- Likelihood of response
- Assess quality of allocation concealment from the assessors for each randomisation
- Willingness to undertake the review process – is this affected by the source of the expert?

Work Plan (Incomplete)

June through December 2013

- Internal consultation, form Steering Committee
- Draft documents required
- Literature Review
- Pilot Study

January 2014 ***through*** *December 2014*

- Conduct definitive study

Outputs

Internal

Report for HTA programme on findings

External

Journal article -> Possible Targets: BMJ Open, Trials, BMC Methodology, BMC Health Services Research, BMC Health Research Policy and Systems

BMJ Open

Involving Clinical Experts in Prioritising Topics for Health Technology Assessment: A Randomised Controlled Trial

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-016104.R1
Article Type:	Research
Date Submitted by the Author:	30-May-2017
Complete List of Authors:	Cook, Andrew; University of Southampton, Wessex Institute; University Hospital Southampton Streit, Elke; University of Southampton, NIHR Evaluation, Trials and Studies Coordinating Centre Davage, Gill; University of Southampton, Wessex Institute
Primary Subject Heading:	Medical publishing and peer review
Secondary Subject Heading:	Evidence based practice, Health services research
Keywords:	peer review, Health Technology Assessment, clinicians

Involving Clinical Experts in Prioritising Topics for Health Technology Assessment: A Randomised Controlled Trial

Andrew Cook [1,2 *], Elke Streit [3], Gill Davage [3]

1 Wessex Institute, University of Southampton, UK
2 Southampton University Hospitals Foundation Trust, UK
3 NIHR Evaluation Trials and Studies Coordinating Centre, University of Southampton, UK

* Corresponding Author – email address andrewc@soton.ac.uk

Clinical Trials Registration: ANZCTR12614000167662 (Retrospectively Registered)
Ethics Approval: University of Southampton, Faculty of Medicine Ethics Committee 8192

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 2 types of review materials (long document vs. short document) and response modes (structured review form vs. free-text email response). Previous and current members of the funder's programme groups were excluded. Response quality was assessed by use of a 4-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 0.335 (Cohen's d = 0.353) over free text. The allocation did not appear to have any effect on the expert's 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: ANZCTR12614000167662

Article Summary

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

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 the right research questions are asked; 2) ensuring that study designs are appropriate and are of methodological quality; 3) ensuring the findings from funded research are available in the public domain; 4) ensuring that funded research is unbiased and usable [1].

The NIHR Health Technology Appraisal (HTA) programme was established in the 1990s, in part to address market failure in UK health research, and is now imbedded in the National Institute for Health Research (NIHR), managed by the NIHR Evaluation, Trials and Studies Coordinating Centre (NETSCC). The programme is the major public funder of pragmatic trials in the United Kingdom, 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 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 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 (or TIDE) panel. These are standing groups of up to twenty clinicians

and lay members, grouped by 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 underway) 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 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 on to the TIDE panel for consideration. Sometimes the secretariat will iterate a point with the clinical expert.

Around thirty percent 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 had to consider what evidence we needed in order to refine the processes which we use to develop the research questions 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 five to ten minutes, 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 or a commissioning brief. The other between being asked to respond using free text, or 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 paper [11] reporting a trial evaluating training for medical students, and 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 two months into our one year recruitment period.

Participants and sample size

The participants were every clinical expert 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 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 size 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 underway, but 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 were not informed of the allocation of participants who provided the responses that they were assessing. However, whether the response was

provided as 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 to 4) 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 the expert and likelihood of responding

We planned to assess whether assessors could identify whether the participants had been sent a vignette or a commissioning brief, to assess the quality of masking.

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 ANOVA, with the quality being the response variable, and the two allocations (vignette vs. commissioning brief, and free text vs. form) as the input variables. Interaction was investigated.

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

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 source of expert on likelihood of response was investigated using chi-squared tests.

All analyses were conducted with R[13].

Sources of data

Data on vignette allocation and quality of responses received was collected specifically for this study. Data on expert's willingness to participate in the reviewing process was 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 to 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.

Firstly 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 updated the inclusion criteria to make this clear.

Secondly, 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 re-randomised. We took the view that if re-randomised, 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 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 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 the commissioning brief; 230 were allocated to a structured response and 230 to free text.

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.

Figure 1 - Consort Diagram of participant flow

[SEE SEPARATE FILE - ParticipantFlowDiagram]

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.

Figure 2 Assessor Scores Across allocated groups

[SEE SEPARATE FILE – AssessorScores]

Counting non-responders as scoring 0 the pooled mean quality score was 2.4, with a standard deviation of 0.95.

The commissioning brief scored 0.037 (Cohen's d = 0.039) extra quality points over the vignette arm; and the structured form response 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 a standard deviation of 0.81. Using data from only responders the commissioning brief scored 0.06 (Cohen's d = 0.071) quality points over the vignette; and the structured response 0.25 (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 pre-defined 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 vs. commissioning brief

Table 1 - Assessment of Masking to Vignette or Commissioning Brief Allocation

Subset	Correct guess of allocation	Incorrect guess of allocation	Significance[1]
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

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 4 sample test for equality of proportions gives a p value of 0.72. We therefore conclude that

[1] Binomial test, assuming that if masking were perfect the correct guess rate would be 0.5, see Statistical methods, above.

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 accepted the invitation. This is a larger figure than the 460 randomised experts, as 42 were approached more than once to review during the course of the study, and only the first acceptance was included in the randomised trial.

We prospectively identified 6 groups of sources from which these experts had been identified.

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 chi-squared test across the whole table has a p-value of less than 0.001, implying a relationship between the source of an expert and their 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 didn't (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 chi-squared 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 2 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 likehood 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 interquartile range of 2 to 4.

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 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 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 is scored a 0, and 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 – firstly to investigate the question around how best to involve clinical experts; but also 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 regards 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 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 providing more information would lead to a superior 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 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 has been able to be undertaken. 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.

Acknowledgements

We gratefully acknowledge constructive advice from our trial steering group (Paula Barratt, Louise Craig, Peter Davidson, Tom Kenny, Sarah Puddicombe, Karen Williams). We thank the NETSCC prioritisation team for facilitating this work, and the panel researcher and consultant advisors who assessed the material receivied from the expert reviewers. We would also like to thank the Southampton University Faculty of Medicine Ethics committee for their advice, the clinical experts who unknowingly contributed to this work, and the referees appointed by the Journal.

Author's Contributions

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.

References

1 Chalmers I, Glasziou PP. Avoidable waste in the production and reporting of research evidence. 2009;**374**:86–9. doi:10.1016/S0140-6736(09)60329-9

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. doi:10.1016/S0140-6736(13)61724-9

4 Health Technology Assessment (HTA) Programme Website. nets.nihr.ac.uk. http://www.nets.nihr.ac.uk/programmes/hta (accessed 17 May2016).

5 Smith R. Classical peer review: an empty gun. *Breast Cancer Res* 2010.

6 Marsh HW, Jayasinghe UW, Bond NW. Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist* 2008;**63**:160–8. doi:10.1037/0003-066X.63.3.160

7 Wessely S. Peer review of grant applications: what do we know? *The Lancet* 1998;**352**:301–5. doi:10.1016/S0140-6736(97)11129-1

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. doi:10.1136/bmjopen-2015-008380

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. doi:10.1186/1741-7015-6-13

10 Hodgson C. How reliable is peer review? An examination of operating grant proposals simultaneously submitted to two similar peer review systems. *Journal of clinical epidemiology* 1997.

11 Hoffmann T, 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. doi:10.1007/s11606-014-2765-5

12 Cook JA, Hislop J, Adewuyi TE, *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**:v–vi–1–175. doi:10.3310/hta18280

13 R Core Team. R: A Language and Environment for Statistical Computing.

14 Cureton EE. Rank-biserial correlation. *Psychometrika* 1956;**21**:287–90.

doi:10.1007/BF02289138

15 Bhurke S, Cook A, Tallant A, *et al.* Using systematic reviews to inform NIHR HTA trial planning and design: a retrospective cohort. *BMC medical research methodology* 2015;**15**:1. doi:10.1186/s12874-015-0102-2

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. doi:10.1186/s12961-015-0025-8

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. doi:10.1136/bmjopen-2012-002521

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* Published Online First: November 2013. doi:10.1186/1745-6215-14-S1-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. doi:10.3310/hta19110

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**:e004121–1. doi:10.1136/bmjopen-2013-004121

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. doi:10.1136/bmjopen-2013-003713

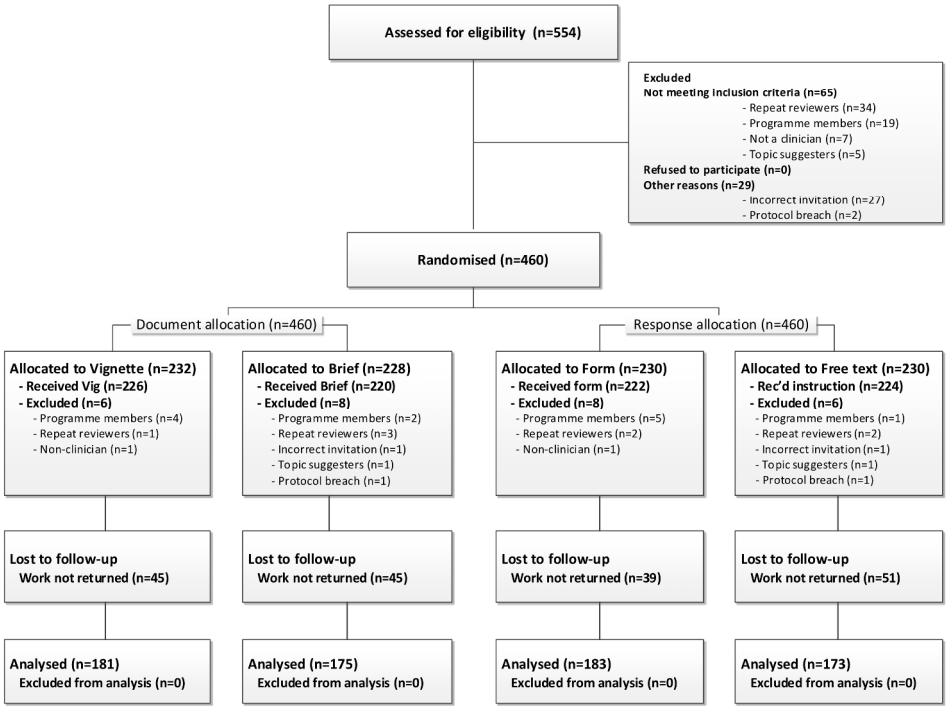
Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors

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 NIHR. Their continuing employment may to some extent depend on the continued funding of NIHR

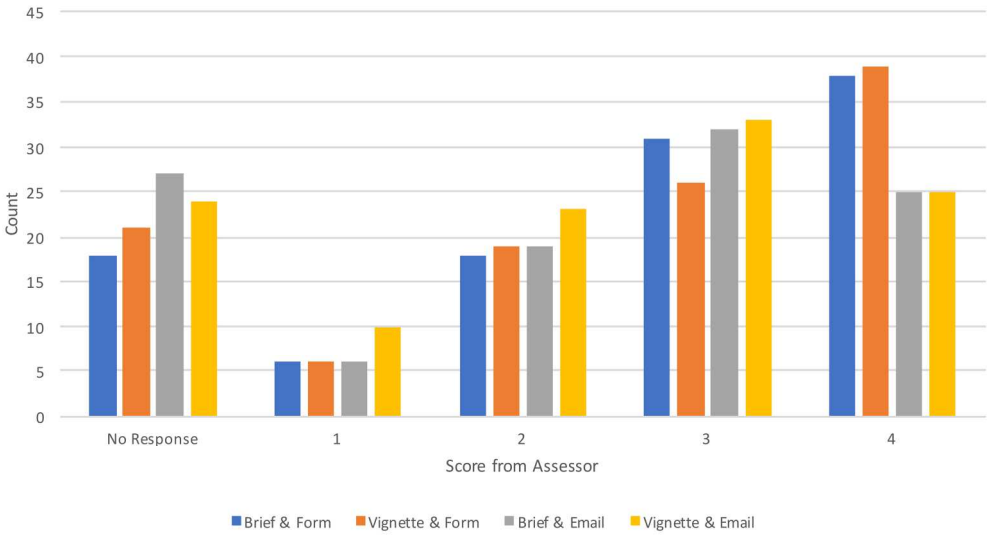
Participant Consent: Not obtained, with the agreement of the University of Southampton Faculty of Medicine Ethics Committee (#8192)

Data Sharing Statement: Anonymised data may be requested from the corresponding author



Consort Diagram of participant flow

277x205mm (300 x 300 DPI)



Assessor Scores Across allocated groups

190x104mm (300 x 300 DPI)

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	4-6
	2b	Specific objectives or hypotheses	6-7
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	7-8
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	10-12
Participants	4a	Eligibility criteria for participants	7-8
	4b	Settings and locations where the data were collected	7-8
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	5-6, 7
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	9
	6b	Any changes to trial outcomes after the trial commenced, with reasons	10-11
Sample size	7a	How sample size was determined	7-8
	7b	When applicable, explanation of any interim analyses and stopping guidelines	10
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	8
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	8
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	8
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	8
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	8

		assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	n/a
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	9, 17
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	n/a
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	12
	13b	For each group, losses and exclusions after randomisation, together with reasons	12
Recruitment	14a	Dates defining the periods of recruitment and follow-up	7-8
	14b	Why the trial ended or was stopped	7-8
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	n/a
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	12, 9
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	12-14
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	n/a
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	12-14
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	n/a
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	18-19
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	18-20
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	18-20
Other information			
Registration	23	Registration number and name of trial registry	1
Protocol	24	Where the full trial protocol can be accessed, if available	By email
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	24

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

Project Protocol

What is the optimum way to approach external clinical vignette reviewers for the NIHR Health Technology Assessment programme: a factorial RCT

12th October 2013
Version 0.4

Change Control

Date	Description of Change	Made By
30th May 2013	First Draft	Andrew Cook
9th July 2013	Following first steering group meeting • factorial design • historical cohort to address invitation wording • Can 'use' assessors be blinded?	Andrew Cook
5th October 2013	Updates to the analysis plan following internal discussion	Andrew Cook
7st October 2013	Updates for external consumption	Andrew Cook
9th October 2013	Changes to expert approach process following first ethics review	Andrew Cook

Abstract

The HTA programme consumes large numbers of clinical experts to inform its commissioning briefs. This takes large amounts of staff time internally, and that of the clinical experts externally.

At a recent induction meeting, two new Panel Chairs suggested that clinical experts do not need an entire vignette in order to make useful and constructive comments - and may be put off by invitation letters which suggest an hour of their time might be needed. He suggested that the 'commissioning brief' portion may be sufficient - and could be processed by the expert in much less time. In addition, if we could offer work which takes less time, we may get a higher uptake rate.

If this is true, it would help optimise the HTA programme's use of clinical experts. We therefore propose a trial of approaches to clinical experts for vignettes using a factorial design to assess two questions

- What is the effect of sending the 'commissioning brief' rather than the vignette to experts for assessment?
- What is the effect of asking experts to respond with a free text response in email, compared to completing a structured form?

Introduction

The NIHR HTA programme commissions research through two main routes:

1. A researcher led route, where investigators propose research questions and methods for answering them to the programme, which then uses its processes to decide which proposals to fund
2. A commissioned route, where the programme identifies research questions of importance to the NHS through a number of processes (literature review, consultations with stakeholders etc) and then asks panels of clinical and lay experts to validate prioritisation decisions. These panels are provided with briefing documents of somewhere between 4 and 10 pages (known internally as vignettes) to help them take these decisions. For those topics which are prioritised a shorter document (known as a commissioning breif) is advertised to the research community. As part of the preparation of these documents external clinical experts are involved to provide information on clinical practice and clinical unknowns.

The HTA programme consumes large numbers of clinical experts to inform its commissioning briefs. This takes large amounts of staff time internally, and that of the clinical experts externally.

Current practice

- Experts are identified
- Experts are approached with offers of work to review the vignette, with an expectation that the task will take around an hour
- The vignette and assessment form are delivered to the expert as email attachments
- Experts are requested to complete the PDF form and return it to the office.

This process represents a fair amount of friction - consumes an hour, working with PDF and attachments - and it is unsurprising that a significant fraction of experts decline to engage.

At a recent induction meeting, two new panel chairs suggested that clinical experts do not need an entire vignette in order to make useful and constructive comments - and may be put off by invitation letters which suggest an hour of their time might be needed, and excessive attachments. He suggested that the 'commissioning brief' portion may be sufficient - and could be processed by the expert in much less time. In addition, if we could offer work which takes less time, we may get a higher uptake rate.

This suggests an alternative process where

- Experts are identified

- Experts are approached with offers of work to review the draft commissioning brief, with an expectation that the task will take around 10 minutes.
- The brief is delivered as the body of an email, with any supplemental questions which the CA or PR wish to ask also in the body of the email.
- The expert responds by replying to the email
- We would offer the vignette if people want it.

This process removes friction, and gives the expert less work, so we might expect more experts to engage with this process. Increasing clinical participation in refereeing would help the programme to demonstrate engagement with the clinical community, but it would only be of use if quality is maintained.

Importance

Expert identification takes up a significant internal resource. If we were able to improve the response rate of experts without decreasing the quality of their input then we may be able to free resource in expert identification. We would also reduce the overall burden on our pool of clinical experts.

Governance

Steering Group

Confirmed Members

- Andrew Cook (CI) - CPHM & Fellow in HTA
- Elke Streit - Panel Researcher
- Peter Davidson - Director of HTA, NETSCC
- Paula Barratt - Senior Research Fellow
- Tom Kenny - Director of External Relations, NETSCC
- Louise Craig - Panel Manager (until December 2013)
- Karen Williams (From December 2013)

Ethics

Will need approval from University of Southampton Faculty of Medicine Ethics Committee

Registration

Will ideally be registered, but no registry yet found which will take an unfunded trial involving neither patients nor a health outcome.

Research Questions

1. What is effect on completion rate and usefulness of responses received when clinical experts in the vignette process are provided with the 'commissioning brief' rather than the vignette for comment.
2. What is the effect on completion rate and usefulness of responses when clinical experts in the vignette process are asked to respond using free text, rather than completing and returning a structured form.

Definitions

Acceptance Rate	The proportion of requests where the work is accepted.
Completion Rate	proportion of experts approached who complete and return the requested work within agreed timescales
Commissioning Brief	The front portion of the vignette, up to and including the 'background box', which will later in the process be used as the basis for the commissioning brief to be advertised.
Usefulness of responses	A decision as to which set of responses is more useful, as assessed jointly by the CA & PR responsible for producing a vignette.

PICOs

Factorial Trial

P	Invited Clinical Experts
I	**1) Commissioning Brief** **2) Email Response**
C	1) Vignette 2) Form
O	a. Return an opinion within time limit b. Usefulness c. Requests for vignette in I1 box **d. Time to respond**

Theoretical Framework

The questions will be investigated using a 2x2 factorial randomised controlled trial Individuals approached to participate will not be informed that other participation options are available (other than those sent the 'brief' will be offered the full vignette on request - they will not however be told there is an alternate arm).

We're interested in assessing how well experts engage with the programme. Hiding the existence of other arms from the experts will prevent them favouring any part of the documentation they receive (such as the 'brief' component) or trying to anticipate what other modes of response might have been offered to them. We believe this is reasonable is the default position for experts is to decline to engage - we will here be testing how different approaches affect that engagement rate - and having those approached know there is an assessment process going on will confound the results.

We do not believe that results for the two allocations - alternate documents, and alternate means of response - will interact, but we will test this as part of the assessment.

Pilot Study

We have conducted a pilot study internal to NETSCC (ie not affecting engagement with people outside NETSCC) to assess

- The record keeping system for randomisation and results
- The practicalities of randomisation

Randomisation

Randomisation will be by precalculated blocks.

We have demonstrated that the blocks can be demonstrated, and that experts can be allocated to a randomisation by an assistant programme manager using a set of work instructions.

We have not yet acted on this randomisation to send any material to experts.

Record Keeping

The main record keeping document for the factorial trial will be a table, with experts recorded on rows, with columns for

- Expert name or ID code
- Document allocation
- Response allocation
- Date of work allocation
- Date of work received (if any)
- Whether the expert requested the vignette

We have demonstrated that

1. An assistant programme manager can enter experts into the table in the order they accept work can be done, in accordance with a set of work instructions
2. The table can be pre-populated with allocation codes
3. The relevant dates can be entered into the table by the appropriate programme manager

Inclusion and Exclusion Criteria

Vignettes

Inclusion

- All vignettes prepared during 2014 rounds.

Except

- Vignettes which ask a methodological, rather than clinical, question.
- Vignettes where the responsible CA does not want external experts.
- Vignettes which have been through previous external review, other than as a panel topic.

Experts

Inclusion

- Clinical experts approached to advise on included vignettes

Except

- Non-clinical experts. Clinical-Academics count as clinical experts. The final decision as to whether an advisor is 'clinical' rests with the Steering Group responsible for the vignette, who may take advice from the relevant CA.
- People who have provided advice on the panel topic from which the vignette was derived
- Current and previous members of HTA prioritisation and TIDE panels

Sampling and Randomisation

Experts will be sampled in line with the usual practice at NETSCC.

Experts who are approached to be offered work will be sent an invitation letter including the text NETSCC is currently assessing what material should be sent to clinical reviewers, therefore please do not be surprised if the material you are sent, or the accompanying instructions, differ from those you may have received in the past.

Within the randomised portion of the study, allocations to each of the two randomisations (document and response) will be pre-generated. Randomisation will be by blocks, with random block sizes of 2, 4 and 6.

Experts will be allocated based on the order their acceptances of work are received, as defined by the time stamp of the acceptance email.

Data Collection

Acceptance and Completion

For the vignette rounds of 2014 we will prospectively collect for each expert approached

1. Panel
2. Vignette ID
3. Vignette Round
4. Was the work accepted
5. Allocated to Vignette or Brief document
6. Allocated to Form or Email response
7. Was the work completed
8. Date work sent
9. Date response received

How useful are the responses?

Each CA & PR team scores each response received on a scale of 0-10. Fractional scores are allowed (eg if a CA score 7 and a PR score 8, the summary result could be 7.5)

Also ask why the grade was given.

Pick list:

- Response ignored
- Told us stuff we knew already
  - From literature
  - From another expert
- Changed the story of the vignette
- Changed the facts
- Changed the science
- Changed the PICO
- Other (specify)

Stopping Rules

Safety

After each vignette round, CA/PR pairs will be asked for each topic

"Was the overall information received from experts to inform your decisions at least as good as you would normally expect".

With a response of 'no' for more than 25% of topics, the steering group will be asked to consider whether the quality of commissioning briefs is at risk, and if so whether the trial should be stopped.

Analysis Plan

Primary Outcome

The main outcome of interest is the usefulness of the response received. This is measured on a scale of 0 to 10, with no response scoring 0. Fractional values are allowed.

The mean and standard deviations of the usefulness score will be calculated

Will be modelled with ANOVA, with usefulness of response being the response variable, and the two trial allocations being the input variables.

We will investigate interaction between the allocations.

Secondary Outcomes

- Likelihood of response
- Assess quality of allocation concealment from the assessors for each randomisation
- Willingness to undertake the review process – is this affected by the source of the expert?

Work Plan (Incomplete)

June through December 2013

- Internal consultation, form Steering Committee
- Draft documents required
- Literature Review
- Pilot Study

January 2014 ***through*** *December 2014*

- Conduct definitive study

Outputs

Internal

Report for HTA programme on findings

External

Journal article -> Possible Targets: BMJ Open, Trials, BMC Methodology, BMC Health Services Research, BMC Health Research Policy and Systems