Author’s response to reviews

Title: Identifying policies and strategies for GP retention in direct patient care in the United Kingdom: a RAND/UCLA Appropriateness Method Panel study

Authors:

Suzanne Richards (s.h.richards@leeds.ac.uk)
Rupa Chilvers (rupa@tangerinebee.com)
Emily Fletcher (e.fletcher@Exeter.ac.uk)
Alex Aylward (aylward@btinternet.com)
Sarah Dean (s.dean@exeter.ac.uk)
Chris Salisbury (c.salisbury@bristol.ac.uk)
John Campbell (john.campbell@exeter.ac.uk)

Version: 1 Date: 09 Jul 2019

Author’s response to reviews:

We thank the reviewers for the feedback on this manuscript. We have summarised our response to review following each comment. The amendments have been track changed in the revised manuscript uploaded.

REVIEWER 1

Comment: Although well written and an important subject, sadly the number of subjects used (9 completed two rounds of Delphi) is insufficient. In a 2007 review on the Delphi process "The Delphi Technique: Making Sense Of Consensus" by Chia-Chien Hsu, The Ohio State University & Brian A. Sandford, Oklahoma State University they stated: "...what constitutes an optimal number of subjects in a Delphi study never reaches a consensus in the literature. Delbecq, Van de Ven, and Gustafson (1975) suggest that ten to fifteen subjects could be sufficient if the background of the Delphi subjects is homogeneous." Making policy suggestions on the basis of the opinions of 9 GPs grading pre-set suggestions is a stretch. I think the authors need to continue the study with more participants (possibly from the second group (national health system experts) before publishing the study.
Our response: We have reviewed the Hsu (2007) paper carefully to see how it might be applied to our paper. Our primary observation is that we did not use the Delphi approach, which is the subject of the Hsu review, but rather the RAND/UCLA appropriateness method (RAM) which has been described as a hybrid of the DELPHI and Nominal Group Technique approaches to consensus building (Humphrey-Murto et al 2016; https://doi.org/10.1080/0142159X.2017.1245856). The RAM approach differs fundamentally from Delphi as it is not intended for idea generation from panel members, but rather starts from the assumption that there are supporting data available to create the initial questionnaire in a highly structured list. Thus the sampling and panel membership requirements differ, reflecting the difference in roles of panellists depending on the consensus building technique selected. The approach we outline follows the methodological guidelines of the RAM method (Reference 20). Here, the RAM developers note that “panels can be of any size that permits sufficient diversity (a minimum of 7), while ensuring that all have a chance to participate (probably a maximum of 15). The best choice within this range depends on the desired geographic and disciplinary representation on the panel”. Our sample size of 9 panellists is therefore consistent with our chosen method. We are also careful to interpret our study conclusions by emphasising that our findings are from the perspective of GP partners (see page 17, paragraph 2 discussion, and page 19 - conclusions) and justify why this is legitimate, but also a potential limitation of the work.

REVIEWER 2

Comment 1: The introduction could be more concise, with a summary of the previous work without going into specific details e.g the numbers of interventions previously tested. It would be enough to state that the evidence of what is effective for GP retention is spare and of low quality. If the specific details are needed, it would be helpful to directly link them to the study objectives and hypotheses.

Our response: We have edited the introduction to make it more concise (see pages 3-5).

Comment 2: The methods are very thorough, but very long (7.5 pages). I needed to read them twice to fully understand. Perhaps a brief overview of the steps at the start would be helpful.

Our response: We are reluctant to add additional sign-posting at the start of the methods section (increasing the length), mainly as the steps reported (sampling considerations, recruitment of panel members etc) follow standard headings. We are similarly reluctant to reduce the content, as a limitation of previous research is that key descriptions of steps, such as panel recruitment, are often poorly reported elsewhere.
Comment 3: There is jargon used that may not be understood by some readers e.g. GP Partners, Medical Performers List.

Our response: We have reviewed the text for clarity. On page 6 we clarify that GP partners are GPs who have leadership and management responsibilities within their practice (which implicitly covers financial investment). The ‘medical performers list’ is already defined (page 7, paragraph 2).

Comment 4: Could you provide more information about the perspectives the different participants brought? E.g. were GP partners contributing the perspective of the average practicing GP? Which of the participants were in positions of decision-making?

Our response: We already state in the methods section (page 6, sampling considerations) that the panel comprised GPs directly responsible for management GP recruitment/retention including GP partners and GPs working in a national role in workforce planning. As we then discuss how panel composition might impact on decision making (strengths and limitations sub-heading, page 17), we feel this issue has already been adequately addressed.

Comment 5: It may be better to describe the creation of the list of strategies first, before the sampling methods and survey administration. I would see the creation of the materials as a different concept from 'Data collection' where it is currently written.

Our response: We have moved the ‘Data collection’ heading so it now appears on page 12 after the description of how the survey was generated, which is now presented under the new heading of ‘Developing the survey’ (page 8). We have also promoted the sub-heading of ‘data analysis’ (page 12), as this was incorrectly presented as subordinate to the heading of ‘Data collection’ procedures.

Comment 6: Pg 12 line 53-55- what was the reasoning behind using a narrower band because these were policy decisions rather than informing clinical decision-making?

Our response: The narrower band for the central zone of ‘uncertainty’ was adopted as this was a more inclusive approach when identifying workforce policies and strategies; this approach is consistent within the RAM guidance (reference 20). When developing consensus for clinical decision making, to avoid direct risks/harms to patients the ‘appropriate’ or ‘inappropriate’ bandings are narrower (i.e. wider central uncertainty banding). This means that fewer statements are endorsed by panellists, making it harder to achieve consensus. In earlier drafts of the
manuscript we had included a fuller rationale for this choice, but removed it to reduce the length of the methods section. We have now reinserted some of this text to clarify this decision.

Comment 7: The reporting of the results was very technical and difficult to extract meaning, whereas I found the discussion much easier to understand and presented information I had not gleaned from reading the results. There is a lot of information for the reader to digest. Consider using some of the descriptions in the Discussion, within the results.

Our response: A key challenge when reporting the output of a RAM study is to succinctly summarise the results, without losing the important detail on the precise statements endorsed or rejected by the panellists. Given the length of this paper, rather than duplicate text in both the results and discussion, we elected to present the lists of statements deemed acceptable/unacceptable, or of uncertain value in the results, providing a more narrative interpretation of the results in the discussion section. We would, of course, be happy to revisit this presentation if the editor felt this to be necessary.

Comment 8: I didn't understand the implications for research and practice. Research has focused on financial, personal and professional support for GPs, yet many of these policy strategies also focused on these issues. I am not sure of the distinction the authors are making.

Our response: We have clarified the text (discussion, page 18) to acknowledge that while we cover the four main areas previously researched, our RAM panel reflected on wider issues within the statements.

REVIEWER 3

We thank the reviewer for their favourable comments regarding our study findings, and of its wider relevance to other European countries facing similar GP retention crises within their workforce.

Comment 3. Keywords: are well selected. Some suggestions concerning key words from MESH to add: considering the methods "Surveys and Questionnaires" and considering the policies "Health Care Reform"

Our response: We have added key words for ‘health care reform’ and ‘consensus method’. We did not include ‘surveys and questionnaires’ as the questionnaire is only used as a tool to elicit consensus.
Comment 4. Errata: Page 6 line 34-35 "In a review of reviews, Misfeldt et al (2014) concluded that that improving the work environment…” to erase one "that".

Our response: We have edited as suggested.

Comment 5. Methods: Page 7 line 35-27 to adapt reference description to the journal style: "to inform UK policy and organisational interventions (e.g. Wright et al., 2008[17]; and Bell et al., 2014[18])".

Our response: We have edited as suggested.

Comment 5, continued. Methods: Authors standardized participants with a common reference document. But it is not clear if they sent the document by e-mail or they did a meeting to explain it. This second option could have been of help to increase response rate. What did they actually do? Did they think about strategies to increase response rate?

Our response: We have added a sentence to explain that participants were provided with contact details for the research team if they needed to discuss the materials they had received electronically (page 12); no face to face meeting took place. It is unknown if a meeting would have increased response rates - our concern was that if study procedures required a meeting, then this might put-off participants as it might mean time away from practice. Thus we opted for online methods, but with the offer of support for those needing it.

Comment 5, continued. Methods: Authors explained how they had adapted RAND/UCLA expert panel method for selection of policies and strategies. A key element of this methodology is to have enough number of answers along the two rounds. Did the authors think of other ways to collect the answers? As phone calls, visits etc…

Our response: An important feature of the RAM approach is that participants are asked to rate statements individually i.e. so that they avoid social desirability bias when generating ratings.

Comment 6. Results: Authors had a small answer rate of 30% (9/28). Did the authors ask the participants what were the reasons for not responding to the questionnaires?

Our response: This response rate is comparable to that observed by Wright et al 2009 (9/32 GPs approached took part). More commonly authors simply omit reporting of the response rates to
invitations to take part by potential panellists (e.g. Tran et al 2008; Bell et al 2014). We have inserted a short paragraph within the discussion referring to the challenge of recruiting panel members (see strengths and limitations sub-heading, page 18), although we would argue that a characteristic of previous research utilising consensus methods is the lack of transparency about the recruitment process.

Comment 6. Results, continued: I have really valued Figure 1. The data collection process for ratings for appropriateness of the 100 statements which clearly shows there has been a positive approach of panellist to reach positive consensus for appropriate statements from 33 to 41. But there is an errata in numbers

Our response: We thank the reviewer for spotting the errata in numbers. After a careful review of figure 1, we have elected to simplify the reporting, so that only the results from round 2 are now presented – as these are the results discussed in the text. All figures presented have been checked against the original data, and we are confident that this figure is now correct.

Comment 6. Results, continued: Concerning table structure on detailed results: it is quite difficult to follow text and tables. Table 2 on methods section shows quite well the questionnaire structure. So I suggest authors to consider that scheme for results in 2 tables: 1) National/Regional or Practice level and 2) GP level and to add 3 columns: mean, implementation model, GP profile.

Our response: During the drafting of this manuscript we have tried many different forms of presenting the results, including that suggested by the reviewer, but when presented this way the tables become very complex, with multiple blank cells as not all sub-groups were tested on each item. We therefore elected to present in a longer list style to aid accessibility of the information, grouping the results within the text. We therefore propose no change to the presentation, as we believe this remains the most accessible way of presenting the data.

Comment 7. Discussion: One clear limitation of the study is the response rate from the expert panel. Although the RAND method is useful with a very small number, the random selection may not add value here as from the sampling finally only have 9 complete rounds answers. It can be of interest to know if professionals who has been invited to participate in the panel, being such a crucial issue, were asked about their refusal to participate: to ask the reason to those who did not want to participate (12/28) and on the other hand, those who, having accepted, did not complete it (3/12).
Our response: We did not capture reasons for declining participation from potential panellists. For those expressing an interest but who subsequently did not complete round 1 (n=2) or round 2 (n=1), we were also unable to collect reasons for their lack of participation as they were non-responders despite reminders being sent. We have made a minor change to the first paragraph of the results section to clarify this (page 13). Although we have inserted a short paragraph within the discussion referring to the challenges of recruiting panel members (strengths and limitations sub-heading, page 18); we reiterate that previous research has not been transparent about the recruitment process.

Comment 7. Discussion, continued: Authors stated: "However, it remains important that the impact of new policies and strategies are evaluated using efficient study designs (e.g. use of routine data and carefully selected performance indicators), and that investigators and policymakers remain alert to potential for both intended and unintended consequences of interventions aimed at maximising GP retention." But there is a simple action to be done, and it is to ask about the reasons to quit to those leaving. Did you consider during the study this survey as a strategic action?

Our response: There is a wealth of qualitative and survey work undertaken exploring the reasons underlying GPs decision to quit, including work already referenced in the text (references 7, 14) undertaken by our team in earlier stages of this mixed method study. There remains potential for unintended consequences arising from new interventions – for example, wider adoption of portfolio careers may improve retention in this cohort of GPs, but have the unintended consequence of driving up work pressure on GPs electing to remain exclusively in direct patient care, increasing quitting.

Comment 8. Conclusion: The study showed some proposal from managers point of view but there is a lack of opinions of those eventually working now: different GP stakeholders (e.g. salaried GPs, or locums) and to include e.g. primary care skills mix.

Our response: We agree that the restriction of panel members is a limitation of this work, and does not reflect the wider composition of the GP workforce in the UK. However, we propose no change as this issue is already discussed carefully in the discussion and conclusions.

Comment 9. List of abbreviations: I suggest to adding the explanation of "GP practices "at risk". It seems at financial risks but it is not clear the meaning for foreign professionals who may read the paper.

Our response: In the results section of the abstract (page 2) we define GP practices ‘at risk’ as those liable to experience GP shortages within the next 5 years, but note that we had not done
this consistently in the methods (Table 2) and results section. Rather than change the abbreviations, we have inserted ‘at risk of GP shortages’ throughout the results section where it occurs (see Tables 2 and 3). We have checked the discussion and note this was fully described on page 16 (first paragraph) and needs no amendment.

Comment 10. Title: I suggest to adding in the title an identifier for country or health system. That is to say: United Kingdom or NHS in the title as there are many particular policies which can only be applied in the NHS and not in other countries or national services.

Our response: We have amended the title to include reference to the UK setting.