Author’s response to reviews

Title: Are family practice trainers and their host practices any better? Comparing practice trainers and non-trainers and their practices

Authors:

Pieter van den Hombergh (P.Hombergh@chello.nl)
Saskia Schalk-Soekar (saskia.soekar@gmail.com)
Anneke Kramer (A.Kramer@elg.umcn.nl)
Ben Bottema (B.Bottema@elg.umcn.nl)
Stephen Campbell (stephen.campbell@manchester.ac.uk)
Jozé Braspenning (J.Braspenning@iq.umcn.nl)

Version: 4 Date: 10 December 2012

Author’s response to reviews: see over
**Point-to-point reply**

**Reviewer 1**: Samantha Scallan  
In my view the study poses some interesting questions regarding the training environment and features that may be indicative of a high quality training environment.

**Major compulsory revisions**

1. I would have welcomed some more information on the ethical considerations presented by the study. The authors state that the study received approval (presumably ethical approval) from Radboud University Nijmegen. I would like to have got a better sense of the consent given by the practices and doctors for the information collected for the practice accreditation program to be used for the research. Were all included and if not how many opted out?  
Based on these remarks and in consultation with the editor, we have changed the statement on the ethical approval. All general practices who volunteered in the Dutch accreditation program agreed that their data could be used for scientific research (anonymous handling). So, we rephrased the sentence on the ethical approval as: “All practices agreed on the use of the data at an aggregated level. Having this kind of informed consent a separate ethical approval is not required under Dutch law.”

2. The authors note that the RCGP is responsible for establishing the standards for training. Whilst the college may have views on this, the GMC is the standard setter: “The GMC is responsible for setting and assuring standards for specialty including GP training leading to award of a Certificate of Completion of Training.” [http://www.gmc-uk.org/education/postgraduate/specialty_including_gp_training.asp](http://www.gmc-uk.org/education/postgraduate/specialty_including_gp_training.asp) Deaneries have developed their own local approaches which reflect the standards set by the GMC, but they differ from deanery to deanery. It is true that it is too strong to say that colleges are responsible for the standards. So we changed the phrasing into: “…have a responsibility in …”.

3. I would have welcomed more detail on the instrument and procedure – who in the practice completes it? How much patient feedback data is captured? How are the various sources of data combined?  
It is fully understandable that you would like to see a more detailed description of the instrument. A full description of the method, the indicators, the validation and the process of data collection has been published elsewhere, see reference 11. We have elaborated in the paper on the procedure. In the paper we reported “It uses a combination of questionnaires that are completed by FPs and staff members, patient questionnaires and observational checklists completed by trained observers.” We added the sentence: “…observational checklists completed by independent external observers during the practice visit. These trained observers collected and processes the data from the questionnaires and the observation in the practice in a database for analysis. For a full description of the method and the process of data collection we refer to a previous publication.” Over the years some items have been changed to adjust the instrument to the ongoing changes in GP-care.

4. As a non-statistician, I would have welcomed a description of the statistical analysis in lay terms. We tried to clarify the analysis. See point 7c, reviewer 2.

5. The implications section is, I feel, under-developed. For instance I would be interested to hear how this research might intersect with work into the quality of educational environment? Based on a recent paper we added: “It also fits into the concern of the FP training institutes to warrant that trainees get the necessary diagnostic and therapeutic skills and see the right patient mix. Giving trainers feedback on the gap in what can be learned in their practice compared to other training practices would benefit the quality of the training.” Ref: de Jong J, Visser MRM, Mohrs J, Wieringa-de Waard M, Opening the black box: the patient mix of GP trainees. BJGP 2011;61 (591):e650-e657.

**Minor Essential Revisions**

N/A

**Discretionary Revisions**

I would have found it helpful to see the VIP scale, perhaps as an appendix. See point 3.
Reviewer 2: Robert McKinley

Comments

Overall: This is an important paper which has demonstrated but may have underestimated the differences between training and non-training practices. It will be of broad interest to educators, managers and clinicians.

1. Aim: To examine the differences [in the structure and process measures] of trainers and non-trainers and training and non-training practices.
   We added the suggestion to the aim.

2. Methods: Numbers: 203 practices (88 non training and 115 training). 138 were either group (n=76) or 2 handed practices (n=62). Do we infer that 65 were single handed practices? How many of these single handed practices were training? How many practitioners worked in training practices and how many non-trainers were there in training practices? This needs to be clarified.
   We clarified the numbers in the method section. “…There were 88 non-training practices (34.9% single-handed) with a total of 164 FPs, and 115 training practices (16.5% single-handed) with a total of 349 FPs.”

3. Data collection: This was a voluntary data collection so presumably there is no issue of selection bias where the scores on the survey affected the training or non-training status. However we are not told by whom the data were collected. Was it self-report completed by the practice/practitioners or by independent external review by trained observers?
   See point 3 from reviewer 1.

4. Measurements: VIP (ref 11) is a 228 (or 208) item schedule for practice assessment but a 227 practice level item questionnaire was used. The origin of the origin of the 142 practitioner level instrument is unclear.
   Small changes have been made in the VIP since 1998 that account for the difference between 228 and 227 items. The original VIP had questions on the practice and the practitioner’s level (ref 11). The number of indicators on the practitioner’s level was 79 in 1998, but we stated the number of items that was 142. See also point 3 from reviewer 1.

5. Analysis: There is likely to be a complex interaction between trainer and practice and trainer and other practitioners in the practices: doctors who are not trainers but working in training practices are likely to have been attracted to working in training practices, having joined the practice have been positively affected by the milieu of the training practice, have been actively preparing to become trainers or have stepped down from being trainers. Therefore it seems odd to treat non-trainers as a homogenous group when an unspecified number of non-trainers work in training practices and are likely to be different from practitioners in non-training practices. Indeed the authors allude to this in the first paragraph of the discussion ‘The presence of a FP trainer in a practice was associated with a positive effect on non-trainers working in the same practice.’
   We agree that it is likely that there is a complex interaction between trainer and practice and trainer and other practitioners in the practices. We treated the non-trainers as a homogenous group, because we focused on the trainers. In this way the maintained differences are more robust. We skipped the sentence in the discussion “The presence of a FP trainer in a practice was associated with a positive effect on non-trainers working in the same practice. Together they work within a much better organized practice than non-trainers”, because our analyses do not allow this statement. We gratefully grasp your suggestion on comparing non-trainers in training and non-training practices for separate paper.

6. It would be useful to compare non-trainers who work in training practices with those in non-training practices: if there was no difference it might suggest that trainers are different from non-trainers and the authors’ assumption may be valid. Alternatively, if they are there are differences between non trainers in training and non-training practices, it is likely that the authors have underestimated the effect sizes and a comparison of trainers vs non-trainers in non-training practices would give a more accurate estimate. There may be more complex ways of controlling for this (perhaps a multilevel analysis with practitioners clustered within practices) but this is beyond my skills.
   See point 5.

7. Definitions:
   a. Urbanisation: how were small village, medium to large town etc defined
      We added the definition in Table 1.
   b. ‘Practice location’: what was the definition of ‘next to the doctors house’ and was it the same for a single and multi-practitioner practice?
“Next to the doctors house” should be taken literally, and it is more common for single-handed practices in The Netherlands. We clarified type of practice and practice location in Table 1.

c. ‘Significantly higher’: (results, first paragraph) and confusion in nomenclature ‘differences between FP trainers and non-trainers that were significant after correction for covariates’: is this effect size or statistical significance and if so at what level?
In the analysis section we stated “Because of the large number of multiple comparisons involved, the effect sizes (Cohen’s d) for each significant difference were calculated.” So for 47 of the 142 items at FP-level we calculated effect sizes; and for 61 of the 227 items at practice level. We described these items (47 and 61) only, if they were still significant after the covariate correction. To clarify this procedure we added to the analysis section, see below.

“Because of the large number of multiple comparisons involved, the effect sizes (Cohen’s d) were only calculated for the significant differences (p<0.05). …… When the covariates (…) were significantly different for the two groups, an analysis of covariance was used (ANCOVA). We considered the significant differences between the two groups only for those items for which an effect size was calculated. We will present effect sizes only for those items that differed significantly (p<0.05) after the covariate analysis.”

8. Results: Table 1: I don’t understand the data for ‘Type of practice’, ‘Practice location’, ‘Urbanization level’: For ‘type of practice’ do the figures represent the average number of practitioners in each type of practice? For ‘practice location’ do the figures represent the distance in Km between the practice and the nearest doctor’s home? For ‘urbanisation’ do the figures represent the means for a scale on which small village is 1 and a large city 4? If so this is not an appropriate way to treat the data.
See point 7a and 7b

9. Discussion: The results of the analysis are appropriately summarised although the statement ‘The presence of a FP trainer in a practice was associated with a positive effect on non-trainers working in the same practice’ is unsupported by the data presented. The authors allude to the complexity of the relationship between having a trainer in a practice and the practice and the other practitioners in the second paragraph.
See point 5.

10. Limitations: (A) The comparison examined structure and process but not patient outcomes. This final leg of the evaluation triad is the most important and its absence should be considered and perhaps how it can be included in the future. (B) The non-training practices volunteered for the study. To what extent could they be better organised practices than the ‘average’ practice thus further diluting the effect size observed
We have elaborated the limitation section on the first point A.
B. We were happy that our sample had both groups of trainers and non-trainers volunteering for the visit. It is certainly true that this could be a positive selection of non-trainers as it is of the trainers. We could reflect on a possible even larger effect if comparing trainers to all non-trainers outside of the accreditation program, but that would only be speculation. No data are available to support this thought.