Author's response to reviews

Title: Key considerations for the experimental training and evaluation of cancer odour detection dogs: lessons learnt from a double-blind, controlled trial

Authors:

Kevin R Elliker (kelliker@hotmail.co.uk)
Barbara A Sommerville (bas13@cam.ac.uk)
Donald M Broom (dmb16@cam.ac.uk)
David E Neal (den22@cam.ac.uk)
Sarah Armstrong (sarah.armstrong@nottingham.ac.uk)
Hywel C Williams (hywel.williams@nottingham.ac.uk)

Version: 2
Date: 12 February 2014

Author's response to reviews: see over
Manuscript: 1340317577102003
Key considerations for the experimental training and evaluation of cancer odour detection dogs: lessons learnt from a double-blind, controlled trial of prostate cancer detection

Responses to Reviewer’s Comments

Dear Prof Kassouf and colleagues,

Thank you for inviting us to submit a revised version of the above paper. We found the reviewers’ comments very thoughtful and helpful, and we thank the referees for taking the time to provide such detailed and useful feedback to help improve the manuscript.

We now deal with each of the two referees’ comments in turn:

REVIEWERS’ COMMENTS

*Please note: all page references refer to the clean (untracked) version unless stated.*

Reviewer #1:

Comment 1.1: The study is well conducted and the statistical analyses are sound. Taking together the hypotheses and the results, I believe that the manuscript is almost ready for publication.

Response 1.1: Thank you for these supportive comments.

Comment 1.2: Figure 1 is not very clear and needs graphical improvement

Response 1.2: We have redrafted Figure 1 using improved, higher resolution graphics.

Comment 1.3: Page 12, line 4: It would be useful in informing the design of future studies to verify how many odours a dog can remember and for how long.” In a recent study on dog’s olfaction it has been shown that when a particular odour is stored in the long-term memory dogs shift their sniffing behavior from the right to the left nostril: this could be extremely useful for “the design of future studies”. Please add and discuss the following paper: Siniscalchi, M., Sasso, R., Pepe, A.M., Dimatteo, S., Vallortigara, G., Quaranta, A. (2011b). Sniffing with right nostril: lateralization of response to odour stimuli by dogs. Anim. Behav. 82, 399-404.

Response 1.3: Thank you for drawing attention to this fascinating paper. We agree that, in future, this line of research could potentially provide a useful measure of whether or not training sample odours are perceived as novel or familiar to the dogs on repeated exposure. We have therefore added a brief discussion of this study on page 12, paragraph 3.

Reviewer #2:

Reviewer 2 provided a number of detailed and helpful comments, which we deal with section by section.
**General comments and abstract**

**Comment 2.1:** In my opinion the authors have done a good job in writing about a experimental that failed and analysing & describing where and why it went wrong. Very valuable lessons to be learned here, and I think it would be worth publishing for the rest of the academic world to learn from, provided some revisions (partly major, partly minor) are made. The field is easily distracted by ‘fantastic’ results and this down- to- earth counterpart is in my opinion at least as valuable.

**Response 2.1:** Thank you for these summative comments. You are absolutely correct in that the field is dominated by “fantastic” results which underscores the publication of “negative” studies in the public domain in order to overcome what is probably a strong publication bias towards “positive” studies in this field.

**Comment 2.2:** In general it’s written very comprehensible, however the phrasing can be more to the point (e.g.: ‘Ideally, dogs should never be presented samples from the same donor on more than one occasion, though in reality it may not be feasible to gain a sufficient number of training samples to do this for many types of cancer. Although pooling of samples appeared to be unsuccessful in the present study in encouraging the dogs to generalise, it would be useful to further explore the chemical evidence for whether or not pooling of urine/biological samples from different donors could assist in creating new headspace odour profiles to widen the pool of training samples.’)

**Response 2.2:** We have reviewed the document and improved the phrasing of a number of paragraphs without changing the meaning. This includes the paragraph provided as an example above (page 14, bullet 1). The changes are easiest to review using the tracked version of the document. As we have made quite a lot of minor changes we have not listed them individually. However, we would be happy to rapidly respond to any questions you might have should any of the changes require further clarification.

**Comment 2.3:** The abstract should undergo major revising and needs to be more to the point and factual, specifically:

a) How many samples? ‘maximum number of training and testing samples available from a large research hospital’ = rather vague and less relevant than the actual amount used.

b) ‘using a study design that also provided information on canine olfactory learning’. Was the analysis of learning part of the design from the start?

c) Results: what are the results? What success rate ‘is expected by chance’ and how far off are the dogs? More factual/ concrete results please rather than interpretation. I could also be worth mentioning that 7 / 10 dogs proved unfit for the job, not because their olfactory capacity was insufficient, but due to ‘character issues’.

d) Conclusion: first state your conclusion; e.g. ‘… the dogs are not able to …’ and then the implications/ recommendations. I would emphasise your study illustrates that is it very easy to draw misleading conclusions about the abilities of dogs to indicate certain odours, unless extremely robust double-blind test trials are conducted.

**Response 2.3:** We have revised the abstract substantially to make it more factual and address the points above:

a) We have added the number of samples used in the abstract methods.
b) The study intended to account for the issue of multiple sampling learning from the outset (e.g. through the use of pooled samples), but we agree the design was not optimised to test hypotheses about this. We agree with the reviewer that, as the study was primarily designed to test hypotheses about the feasibility of detecting a cancer odour, the clarity of the aims/methods is improved by removing reference to investigating multiple sample learning as a primary goal. We have therefore deleted the sentence of concern (Methods section of abstract).

c) We have added the specific results (specificity and sensitivity) and the finding that 7/10 dogs were unsuitable in the abstract results.

d) We have re-written the abstract conclusion to clarify the main conclusions and emphasise the point made by the referee above.

Comments on Background section

Comment 2.4: relatively ease = relative ease?
Response 2.4: We have corrected this typographical error (page 4, line 2).

Comment 2.4: Last paragraph: long sentence; without reading the rest of the article it’s a bit hazy what you actually mean to do
Response 2.4: We have improved the clarity of the paragraph (page 5, paragraph 2) and removed reference to multiple sample learning (see Comment 2.3b).

Comment 2.5: What the difference between: ‘pseudoreplication of odour samples from the same donors during analyses’ and ‘samples re-used multiple times during testing, potentially making them familiar to the dog’? using donors twice vs using samples twice?
Response 2.5: We agree these are, in effect, the same issue, and have therefore combined them into one point (page 4, paragraph 3)

Comment 2.6: A lot of emphasis is made on the learning aspect being a major research goal. I can imagine that this was not necessarily so on the start of the project. I can imagine that when the experiment did not succeed, the researchers investigated what went wrong, which is of course very commendable. However, in this paper it is presented like this was the idea from the start and the set-up was especially designed to do this. I wonder: what aspect in the setup was especially designed to test whether ‘…multiple sample learning may impact upon the training of dogs.’ I would either specify this, or otherwise present your research as it is, which interesting enough whether or not you set out to analyse where it went wrong from the start.
Response 2.6: As discussed in Response 2.3b, we agree that the clarity of the paper is improved by removing reference to investigating multiple sample learning as a primary goal. We have therefore removed reference to this in the Background (page 5, paragraph 2).

Comments on Methods section
**Comment 2.7:** Very commendable that ten different dogs were used. However, since the majority of the dogs proved to be ‘untrainable’ I wonder how they were selected? Is there an explanation that so many were not trainable?

**Response 2.7:** We agree that the low number of dogs that passed the initial stages of training was a significant limitation. A wide range of factors affects whether dogs are suitable for specialist working tasks and high failure rates for dogs undergoing training for working roles are common (two examples have been added to the manuscript reporting 50% and 70% failure rates).

We have clarified in the Methods (page 5, paragraph 3) that dogs were initially recruited for involvement in the study based on the opinion of professional dog trainers or behavioural scientists, but then underwent more detailed selection based on the dogs’ ability to detect odours in the initial training stages.

We have also added a section discussing why only three dogs passed initial training and the limitations due to this in the Discussion (page 13, paragraph 4).

**Comment 2.8:** Impressive to have such a large number of donors

**Response 2.8:** Thank you – we see this as one of the major strengths of the study compared with other studies in this area, thanks to the involvement of an NIHR senior investigator with a research interest in prostate cancer.

**Comment 2.9:** For how long were the samples stored? More than a few months? Despite rigorous samples techniques the odour could be influenced by ‘aging’, even on -20°C.

**Response 2.9:** Most of the samples were presented to the dogs within 60 days though some were presented up to 6 months after initial collection. We have added a note to this effect to the Urine sample collection and preparation section (page 6, paragraph 3). It is possible odour decay on storage may have been a risk, though similar storage durations were found to be successful in a study on bladder cancer (reference 8 in manuscript), and in that study even air-dried samples were successfully used.

**Comment 2.10:** Were the samples taken after the diagnosis had been made, so could it be the invasive procedure (e.g. taking a biopsy and therefore ‘opening’ the prostate or tumor) that leaves the smell? I take it that the healthy donors had not undergone any invasive procedure; how many healthy donors were there (compared to BPH controls?) How sure are we that the BPH controls indeed did not have cancer; were biopsies taken in all of them? And again were the samples taken before this procedure?

**Response 2.10:** This is a good point and a general concern in all studies involving a diagnostic biomarker where you have one group (the cases) who have the disease and the other group who either do not have the disease at all, or are at low risk.

We have clarified in the methods (page 5, line 30) that all our positive samples were from men with prostate cancer previously confirmed by biopsy. We have also clarified (page 6, paragraph 1) that 13 control samples used for training were from men who had also undergone prostate biopsy (results negative). It is therefore unlikely that an odour associated with the invasive
procedure *per se* would have affected the results, as a proportion of controls had also undergone the same invasive procedure.

We have clarified that 10 control donors were apparently healthy men without clinical symptoms of BPH (page 6, paragraph 1).

As discussed in the Discussion (page 13, paragraph 2) control samples were mainly men with a PSA <0.5 ng/ml, who have a very low risk of ever getting clinical prostate cancer over their lifetime, so were the best available controls available to us (accounting for age).

**Comment 2.11:** Design seems to be very adequate, although ideally more positive samples would be available, but as rightly stated it’s not that easy to get a larger number of samples. I expect wide CI on these numbers.

**Response 2.11:** Thank you for the supportive comments on the robustness of the design. We agree regarding the availability of positive samples but, as discussed in the paper, this is likely to be a limitation of all studies of this type.

**Comment 2.12:** Under ‘training procedure’, 2nd paragraph: because should be became?

**Response 2.12:** Thank you for spotting this – we have corrected it.

**Comments on Results section**

**Comment 2.13:** Where are the patient characteristics on the CaP patients? E.g. PSA and Gleason score? Have they got severe metastasised cancer or small, relatively innocent tumours?

**Response 2.13:** We have added Table 1 which provides details of the samples used in Tests 2 and 3, including PSA and Gleason scores (page 19). We have also clarified in the Urine sample collection and preparation section of the methods that the degree of disease in CaP donors varied from small, relatively innocent tumours to metastasised cancer (page 5, paragraph 5).

**Comment 2.14:** I do not understand the discrepancy between the stage 3 ‘new samples’ and the formal testing results. Was the dog training with the samples (during stage 3) before the result was scored?

**Response 2.14:** The stage 3 training was run handler-blind (single-blind) but not double-blind as the testing stages were (which is why they were required). We have offered two explanations in the manuscript:

- It may have been a statistical artefact due to the small number of donor CaP samples (hence the need for tests to be conducted using urine samples from a larger number of donors), mentioned in the results section (page 10, paragraph 2).

- The dog involved may have had some ability to detect prostate cancer, but this was disrupted during the formal tests due to the different rewarding procedure used to allow double-blinding (discussed in Rewarding techniques during double-blind trials, page 12 paragraph 4).
Comment 2.15: The fact that only 3 / 10 dogs proved trainable is remarkable. Again, how were they selected? In my opinion this is one of the major drawbacks of the study and it should be made more clear (also in the discussion) that it’s not so much so that none of the 10 dogs were capable, but none of the 3 dogs were capable.

Response 2.15: We agree and, as discussed in response 2.7, we have clarified the selection of dogs in the methods and discussion.

Discussion:
Comment 2.16: Our study illustrates that it is very easy to draw misleading conclusions about the abilities of dogs to indicate certain odours, unless extremely robust double-blind test trials are conducted. I think this should be your (one of your) most important conclusion(s), and should receive more emphasis (e.g. in the abstract).

Response 2.16: Good idea - we have now added this point to the abstract and conclusion.

Comment 2.17: Recommendation 4: double BLIND instead of double bind

Response 2.17: Thank you for spotting this – you caught us in a bind which we have now corrected.

Other changes
Please note that we have removed reference to Stage 1 (food training) as a formal training stage in order to make the abstract and results more concise, as this was not a significant part of the study. The use of food to initially encourage dogs is still mentioned in the methods. Training Stages 2 and 3 have therefore been renamed Stages 1 and 2 in the revised manuscript.

Thank you again for reviewing our work for potential publication. We hope these responses have satisfactorily addressed all of the comments made. Please don’t hesitate to contact us if we can provide further clarification on any of the points.

Here is a list of what we have resubmitted:

1. This cover letter
2. One clean copy of our revised manuscript (includes tables)
3. One tracked changes copy of our revised manuscript (includes tables)
4. Figure 1 (revised)

None of the author’s conflicts have changed since the original submission.

We look forward to hearing from you.

Yours sincerely

Hywel C Williams DSc
Centre of Evidence-Based Dermatology,
University of Nottingham
Tel: 0115 82 31048
Fax: 0115 82 31046