PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Clinical decision making in a high risk primary care environment: a qualitative study in the UK
AUTHORS	John Balla, Carl Heneghan, Matthew Thompson and Margaret Balla

VERSION 1 - REVIEW

REVIEWER	Erik Stolper, GP, PhD. Maastricht University. Faculty of Health, Medicine and Life Sciences, CAPHRI School for Public Health and Primary Care. Department of General Practice. PO Box 616 6200 MD Maastricht The Netherlands
REVIEW RETURNED	26/10/2011

RESULTS & CONCLUSIONS	The topic of the research, clinical reasoning in OOH care, is very interesting and I am happy that the authors started to study this specific domain. They did a lot of work and their results are intriguing. Thanks for their efforts. However, there are in my opinion some problematic questions. 1) Several times the authors refer to their in an earlier publication described model which is the theoretical basis for their approach to the topic of research (e.g. 'strong theoretical framework', 'derived from the dual theory of cognition'). For the usual reader, it is not clear at all what this framework meant for this study and in the discussion it remains unclear too. There is no obvious relationship between this strong theoretical framework and the methods, the results or in the discussion, at least it is not described in their manuscript. 2) I knew their earlier publication and I read it again. In their approach the non-analytic reasoning process is similar to pattern recognition and that is right but there is more. Recognition can be considered as characteristic for intuition (see Kahneman&Klein 2009) but intuition is more than pattern recognition. The role of intuition can be seen in diagnostic gut feelings too. Several reports of research into the diagnostic role of gut feelings in primary care have been published in the past few years. 3) I am curious to the questions of the semi-structured interviews. Please add all these questions to this manuscript. I assume that these questions have been derived from the model. If there had not been a question about the role of intuition, the authors might have missed information about the significance of gut feelings in clinical reasoning in OOH care.

REVIEWER	Marloes A van Bokhoven, MD PhD GP and lecturer Maastricht University CAPHRI School for Public Health and Primary Care

	Department of General Practice Maastricht, Netherlands I have not competing interests
REVIEW RETURNED	28/10/2011

THE STUDY	Desciption of participants: The authors do describe the selection procedure and a few characteristics of the participants. However, they do not justify why they aimed for a convenience sample instead of a purposive sample. As the aim of qualitative research is to ensure that all available opinions are represented in the results it is in my opinion necessary that the authors comment whether or not they succeeded to reach this aim. They also do not mention which participant characteristics they expect to influence divergent opinions. Although the authors do state that sampling is a limitation of the study (page 9), they explain this by stating that the sampling was not random, which, as mention above, is not the main reason that sampling is not the strongest point of the study. abstract: In my opinion the recommendations are too strong given the exploratory, qualitative character of this study. These can be found both in key messages, abstract, discussion and conclusions (see also below) References: The authors did a broad study with three themes: the context of out of hours care (OOH), diagnostic reasoning in OOH and recommendations for GP-trainees. They only present 15 references of which, as far as I can see, 8 do not refer to peer reviewed papers. I see, for example, only 1 reference to the extensive body of literature about clinical/diagnostic reasoning and none to the literature on the development of clinical expertise. Also, references to the non-cognitive aspects of diagnotic reasoning are lacking (e.g. Stolper et al but also the work of the authors' own group in collaboration with Leuven (Van den Bruel). Finally, the authors give several recommendations for further quality improvement interventions and teaching but they do not refer to literature about these topics.
RESULTS & CONCLUSIONS	Results: In my opinion the results section is rather short. I would like to suggest that the parts of the discussion section that present new results (e.g. page 12, lines 10-14; lines 31-37) are transferred to the results section. I did not find an interview scheme for the semi-structured interviews. In my opinion this would be helpful for the interpretation of the results. The results are somewhat superficial, but this may be due to the exploratory character of the study. In my opinion the contrast between results and interpretation are not always clear. In my opinion it would be better if presentation of results and interpretation by the researchers are described in the results section. Discussion about generalisation and implication can then be described in the discussion section. There is a limited number of quotes given. May I suggest that the authors use different quotes in the text than in the text boxes? Currently it is not always clear how the quote reflects the the concepts e.g. box 1a: fire fighting, 2nd quote; box 1b: special problems: it is not clear why psychiatry is different during OOH; The authors did not describe the issue of triangulation.

	Conclusions: My main concern with this manuscript is that, although I do find the suggestions regarding feedback and training of GP trainees very interesting, I think it is too early to draw these conclusions from these results. The study is a broad, qualitative, not very in-depth study. In my opinion several steps are now needed, partially mentioned by the authors in the 'further research' section: triangulation, quantitative research and then translation of the barriers and facilitators into interventions. Also, the authors elaborate on the benefits of suggested interventions. However, practice is more complicated than the authors suggest and the suggestions are not derived sufficiently from the data. Therefore, I would suggest that the authors keep this discussion out of the manuscript. E.g: page 14 on computerised feedback. In my OOH clinic we have a feedback opportunity but its use in practice is more complicated and time consuming than the authors suggest. In addition, the authors suggest that the feedback can be used as indicators of functioning. First, the possibility of 'punishing' employees based on the feedback data negatively influence the aim of quality improvement, as has been described with blame free reporting of incidents. Secondly, the use of feedback/indicators for external purposes such as judgement of functioning demands much more from the quality of the feedback/indicator than when you use it for personal feedback.
GENERAL COMMENTS	Though my comments may seem rather critical I would like to state that I think the authors studied a relevant, important and interesting topic. I hope that they will find the means to proceed with research on this topic. Unfortunately, the review form is better suited for quatitative studies.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: we are grateful to the first reviewer for his thoughtful comments which help us to clarify a number of issues raised in our manuscript.

1.1. Clarification of the relationship between the theoretical framework with the methods, results and discussion: we agree that clarification of these relationships is necessary. We deal with this in the section on methodology, sub-section Design, paragraph 2, appearing in red script. We further elaborate on this issue under Discussion paragraph 3.

2. 1. The role of recognition in intuition: we are familiar with the work of Kahneman and Klein (2009) and we refer to them as Reference 10.

2.2. The role of diagnostic gut feelings to intuition and their role in non-analytic reasoning: we are familiar with the work of the first reviewer and his colleagues on diagnostic gut feeling and refer to them as Reference 11.

3.1. Addition of the list of questions used in the interview and their relevance to the theoretical model: We added Appendix 1: List of interview questions and their relevance to the theoretical model.

3.2. Clarification of questions related to intuition: As indicated under the section of ‘Strengths and limitations’: We designed the study to allow us to focus on specified steps of the clinical reasoning process, e.g. closure. This means that some steps could not have full coverage of issues, such as some of the crucial intuitive aspects of the early stages of diagnosis. This will require further qualitative work.

Reviewer 2: we are grateful to the second reviewer for her thoughtful comments which help us to clarify a number of issues raised in our manuscript.

1. Description of participants

1.1. Justify why aimed for a convenience sample instead of a purposive sample: We explain our reasoning under ‘Setting and cohort’: In our experience and in reports in the literature, busy professionals are relatively difficult to persuade to find time for an interview. Therefore one is generally unlikely to find random cohorts, the participants coming through word of mouth and professional contacts. The only variable we would expect to have an impact for this study was length of experience and in this respect our cohort was relatively well represented.

1.2. As the aim of qualitative research is to ensure that all available opinions are represented in the results, the authors comment whether or not this succeeded to reach this aim: Details of our limitations are explained as above under Methods. We believe that we achieved our expressed aims.

1.3. Not mentioned which participant characteristics expected to influence divergent opinions: As under ‘Setting and cohort’:The only variable we would expect to have an impact for this study was length of experience and in this respect our cohort was relatively well represented.

1.4. Although the authors do state that sampling is a limitation of the study (page 9), they explain this by stating that the sampling was not random, which, as mention above, is not the main reason that sampling is not the strongest point of the study: as above and Methods, we provide reasons for our sampling methods: We chose 30 minutes to minimise the pressure on busy work schedules. From past experience, most GPs would be interviewed before or after consulting and often during their brief lunch breaks. The time allocated was deemed sufficient to capture the data on the issues on which we were focused.

2. Abstract:
Recommendations are too strong given the exploratory, qualitative character of this study: We have revised the language used in recommendations as highlighted in the Abstract.

3. References: The authors did a broad study with three themes: the context of out of hours care (OOH), diagnostic reasoning in OOH and recommendations for GP-trainees.

3.1. They only present 15 references of which, as far as I can see, 8 do not refer to peer reviewed papers. I see, for example, only 1 reference to the extensive body of literature about clinical/diagnostic reasoning and none to the literature on the development of clinical expertise: We added a further 8 references as shown in the List of References.

3.4. References to the non-cognitive aspects of diagnotic reasoning are lacking (e.g. Stolper et al but also the work of the authors' own group in collaboration with Leuven (Van den Bruel): We have added a reference by Stolper et al and there is also a reference to Gabbay.

3.5. They give several recommendations for further quality improvement interventions and teaching but do not refer to literature about these topics: We referred to Biggs and Tang a foremost authority on education and also added Ericsson. Our response is detailed under the section on ‘Implications for practice’.

4. Results:
4.1. In my opinion the results section is rather short. I would like to suggest that the parts of the discussion section that present new results (e.g. page 12, lines 10-14; lines 31-37) are transferred to

the results section: In response to these and subsequent comments we have significantly rewritten the Discussion, paragraph 3. We draw together threads of the results and theoretical base and we feel that in this new format it seems appropriate to leave these lines in the discussion.

4.2. I did not find an interview scheme for the semi-structured interviews. In my opinion this would be helpful for the interpretation of the results: We added Appendix 1.

4.3. The results are somewhat superficial, but this may be due to the exploratory character of the study. In my opinion the contrast between results and interpretation are not always clear: We believe that our response, as in 4.1. above, deals with this issue.

4.4. In my opinion it would be better if presentation of results and interpretation by the researchers are described in the results section. Discussion about generalisation and implication can then be described in the discussion section: see comments under 4.1 above.

4.5. There are a limited number of quotes given. May I suggest that the authors use different quotes in the text than in the text boxes? We eliminated repetition using new quotes, as in red in the Boxes.

4.6. Currently it is not always clear how the quote reflects the concepts e.g. box 1a: fire fighting, 2nd quote; box 1b: special problems: it is not clear why psychiatry is different during OOH: Box1a: quote changed. Box 1b: quote deleted.

4.6. The authors did not describe the issue of triangulation: We deal with this issue under ‘Limitations’: These qualitative studies lead to the development of inventories suitable for triangulation through quantitative studies. Also, under ‘future studies’:
We also propose to develop a quantitative questionnaire based on our qualitative studies and on leads that we find in the literature. Such studies, extending beyond our cohort, may then provide data to allow greater generalisability of the results.

5. Conclusions: Our responses to these comments are detailed in ‘Implications for practice’ and also highlighted in Conclusions.

5.1. My main concern with this manuscript is that, although I do find the suggestions regarding feedback and training of GP trainees very interesting, I think it is too early to draw these conclusions from these results: Recognising the limitations of our study, we also find that our conclusions are consistent with the literature on the importance of feedback on performance and settings for fruitful reflection on experience. We believe that given this congruence between the relevant literature and our proposals, our recommendations merit cautious acceptance and evaluation.

5.2. The study is a broad, qualitative, not very in-depth study. In my opinion several steps are now needed, partially mentioned by the authors in the 'further research' section: triangulation, quantitative research and then translation of the barriers and facilitators into interventions: As 5.1. above and also comments in 4.6. above.

5.3. The authors elaborate on the benefits of suggested interventions: As under ‘Implications for practice’: To this stage excellent improvements have been demonstrated in interventional specialties but not in primary care. However, in view of the confluence of the principles of deliberate practice and deep learning approaches, we recommend the evaluation of deliberate practice training models for trainees.

5.4. Practice is more complicated than the authors suggest and the suggestions are not derived sufficiently from the data: These systems need to be implemented with sensitivity, recognising the risk

of creating anxiety and dislike of being monitored. Feedback needs to be seen as non-judgmental and to be used by the recipient and trusted colleagues only. The importance of regular feedback on performance to gain expertise is well accepted in the literature.

5.5. Therefore, I would suggest that the authors keep this discussion out of the manuscript. E.g: page 14 on computerised feedback. In my OOH clinic we have a feedback opportunity but its use in practice is more complicated and time consuming than the authors suggest: as in 5.4. above.

5.6. In addition, the authors suggest that the feedback can be used as indicators of functioning. First, the possibility of 'punishing' employees based on the feedback data negatively influence the aim of quality improvement, as has been described with blame free reporting of incidents. Secondly, the use of feedback/indicators for external purposes such as judgement of functioning demands much more from the quality of the feedback/indicator than when you use it for personal feedback: In line with these comments we altered our recommendations as 5.4.

VERSION 2 – REVIEW

REVIEWER	C.F.Stolper, MD, family physician, PhD. Maastricht University. Faculty of Health, Medicine and Life Sciences, CAPHRI School for Public Health and Primary Care. Department of Family Medicine. P.O. Box 616 6200 MD Maastricht the Netherlands.
REVIEW RETURNED	21/11/2011

The reviewer filled out the checklist but made no further comments.

REVIEWER	dr. MA van Bokhoven, MD PhD GP and university lecturer Maastricht University, CAPHRI School for public health and primary care, dept of General Practice Maastricht The Netherlands
REVIEW RETURNED	25/11/2011

THE STUDY	1. I am still not satisfied with the authors' response regarding the sampling. Probably I did not make myself clear enough but the authors give me the impression that they used a sampling strategy that applies to quantitative research not qualitative research. In their response they state for example: 'Therefore one is generally unlikely to find random cohorts'. My message is that the authors should not aim for a representative sample but for a purposive sample instead. This implies that not experienced GPs 'was relatively well represented' but that the authors should have aimed for GPs with divergent opinions in their sample. If that was not feasible (though the study period was short with only 2 months duration), the authors should at least report if they think all different opinions have come up in their study. I do not read if inexperienced GPs responded differently and I am also not convinced that GP experience is the only factor that influences the GPs' opinions. e.g. location of OOH practices (low SES area or high SES), differences in previous training in emergency medicine of GPs can influence their opinions. Even if the sample is a convenience sample, the authors should

	report if GPs with these characteristics were included. 2. the authors still give the impression in their abstract that they studied the effect of feedback on the quality of care by stating that they study supports that ; improvements could be achieved...'. That is not what their study was about. Similar to their conlusion that trainee support is required, while theyr only asked GPs what they would recommend to their trainees. I would like to thank the authors for presenting their interview scheme, which clarifies a lot. However, in my opinion the questions influenced the results several times through cueing. E.g. the authors raised the issue of safety netting themselves. In my opinion it is then no longer legitimate to state that safety netting is an important issue.
RESULTS & CONCLUSIONS	I am very sorry, but I am still not convinced that the conclusions are sufficiently derived from the data. E.g. several participants are reported to descibe the OOH setting as a place where you do not often get feedback on your actions. The authors conlude from this that 'an automated, regular and timely system of feedback to GPs and other clinical staff practising in the OOH setting is essential'. They studied their topic in 1 British area and in general practice only. In my opinion the conclusion that the results can be generalized to all general practice and specialty care as well is not correct without further study. The effect of different quality improvement strategies depends on the context very much. E.g. at emergency departments there are many diagnostic facilities that are sometimes routinely used. This influences OOH diagnostic reasoning (or even prevents reasoning as complaints may be treated according to a protocol). Although I am happy that the authors added several new references still 10/23 refer to books. Although I realise that the authors, due to the very broad character of their study, could find relevant studies in many different areas, I would have preferred reviews from peer reviewed journals over book chapters.

VERSION 2 – AUTHOR RESPONSE

Responses:

1. Sampling
1.1. aimed for GPs with divergent opinions in their sample. If that was not feasible (though the study period was short with only 2 months duration), the authors should at least report if they think all different opinions have come up in their study: we believe we answered this question under ‘Limitations’:We designed the study to allow us to focus on specified steps of the clinical reasoning process, e.g. closure. This means that some steps could not have full coverage of issues, such as some of the crucial intuitive aspects of the early stages of diagnosis. This will require further qualitative work. i.e. all different opinions did not come up, but we have a spread with regards to the questions we asked.

1.2. e.g. location of OOH I do not read if inexperienced GPs responded differently and I am also not convinced that GP experience is the only factor that influences the GPs' opinions. practices (low SES area or high SES), differences in previous training in emergency medicine of GPs can influence their opinions. Even if the sample is a convenience sample, the authors should report if GPs with these characteristics were included: we stated under ‘Design’: The only variable we would expect to have an impact for this study was length of experience and in this respect our cohort was relatively well represented. This is based on the data, where a number of GPs told us how they changed with experience. We agree there are multiple other influences, which is why further studies are needed, as we state in the manuscript. We feel the cohort is too small to further sub-categorise our findings or speculate on causality.

2.
2.1 Abstract: impression in their abstract that they studied the effect of feedback on the quality of care by stating that they study supports that; improvements could be achieved: we stated that “findings support suggestions” re feedback. Nowhere do we say we studied the impact of feedback on performance. What we say is that our findings derived from our data support what our GPs said and others in the literature have said.

2.2. conlusion that trainee support is required, while theyr only asked GPs what they would recommend to their trainees: we made it clear under ‘Discussion’ that all our conclusions are based on our data, i.e. what the GPs told us and the relevant literature on the subject, consisting of previous reports and the theories of education we also refer to.

2.3. interview scheme: questions influenced the results several times through cueing. E.g. the authors raised the issue of safety netting themselves. In my opinion it is then no longer legitimate to state that safety netting is an important issue: the question we posed was: Q4.4: Can you tell me how safety netting comes into this? This came as a clarifying question after the GPs described what they did and not before. They then stated that, yes what they did was safety netting but they were not naming it as such. Some of them were then also critical about shortcomings of what they had done. We never asked it as a leading question, but at times as a clarifying question in some cases.

2.4. not convinced that the conclusions are sufficiently derived from the data: Under ‘Conclusions’ we state: Our study
provides further support to the literature on reflective practice and educational programs in clinical settings. We suggest that improving feedback to GPs about their clinical decisions and providing opportunities for reflection on OOH practice may be valuable for ongoing review and improvement of clinical practice. Our data derived from interviews supports the literature. The conclusions are supported by literature, but need evaluation, as we state below.

2.4.1. to descibe the OOH setting as a place where you do not often get feedback on your actions. The authors conlude from this that 'an automated, regular and timely system of feedback to GPs and other clinical staff practising in the OOH setting is essential': Under implications for practice we state: “Recognising the limitations of our study, we also find that our conclusions are consistent with the literature on the importance of feedback on performance and settings for fruitful reflection on experience. We believe that given this congruence between the relevant literature and our proposals, our recommendations merit cautious acceptance and evaluation.” Further on we also state: “we recommend the evaluation of deliberate practice training models for trainees.” We repeatedly stress the need for evaluation of the interventions we recommend.

2.4.2. They studied their topic in 1 British area and in general practice only. In my opinion the conclusion that the results can be generalized to all general practice and specialty care as well is not correct without further study: Under ‘Limitations’ we state: “We have not performed similar patient based studies in other settings in the UK or in countries with different health systems and to this stage the research has not included other specialty groups. Further down we state: “On account of the strong theoretical base, our findings, albeit with different emphases, are likely to have some degree of generalisability to other specialties and health systems.” At no stage do we say or imply that we conclude that: ‘the results can be generalized to all general practice and specialty care as well’.

2.4.3. The effect of different quality improvement strategies depends on the context very much. E.g. at emergency departments there are many diagnostic facilities that are sometimes routinely used. This influences OOH diagnostic reasoning (or even prevents reasoning as complaints may be treated according to a protocol): We completely agree with the above. This is why we stress that our data

come from a single region and further studies are needed in different parts of the UK. However, as stated, other UK reports come up with similar trends.

2.4.4. refer to books: preferred reviews from peer reviewed journals over book chapters: the books we refer to are written by recognised authorities on their subjects. They provide a well-researched overview in each case, particularly suitable for a general readership.