

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form ([see an example](#)) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Identifying early warning signs for diagnostic errors in primary care: a qualitative study
AUTHORS	Balla, John ; Heneghan, Carl; Goyder, Clare; Thompson, Matthew

VERSION 1 - REVIEW

REVIEWER	Frank Buntinx Professor Dept of General Practice University of Leuven (Belgium) & Maastricht University(The Netherlands)
REVIEW RETURNED	30-Jun-2012

THE STUDY	Q 7, 9, 10: not applicable to a qualitative study report. Last Q: The checklist states that saturation was "reached at about 2/3 of way" Unless I missed it, this does not appear in the manuscript.
REPORTING & ETHICS	Q2: Formally, it may be appropriate to state that all 15 GPs gave informed consent to use their data.
GENERAL COMMENTS	<p>General remarks: Relevant and nice paper. Appropriate methods. Sometimes difficult reading (I indicated the most difficult part, but to a lesser extent this is the case for the whole manuscript).</p> <p>Minor remarks: Introduction: This section is quite long. The paper will become more readable if it is seriously shortened. The sequence of the terms differs in line 1-3 from line 4. Methods: Data saturation is not mentioned in the paper (the information is available in the checklist). § 2: Although (as a non-native English speaker) I suppose that the language is correct, especially this paragraph is difficult reading. It will be difficult to understand by non-English readers. I think that it would be possible to reformulate this in more simple and direct language. Results: Line 3-5 from the end (Do I want ... mayhem): I don't understand this sentence. Discussion: I did not find a figure 1! Conclusion: Do you really consider this section necessary? I would suggest to drop it. Nothing new is presented. Tables,: Table 1 / second column: your message would be clearer if you add to each sentence something like "instead of ..." with the correct</p>

	<p>diagnosis. Box 2: Reconstruction / Presentation, 1th §: What is PR? When were PR and bloods performed? Six months ago or now? Analysis / Presentation / second §: Are you sure you didn't switch around rule in and out?</p>
--	--

REVIEWER	<p>Dr Lucy Brindle Improving Earlier Diagnosis Research Group Lead/Lecturer University of Southampton UK</p>
REVIEW RETURNED	08-Jul-2012

THE STUDY	<p>It is well recognised that retrospective interviews are an unreliable method for capturing what actually went on within an interaction or institutional practice. Whereas interviews might have an important role in exploring cognitive processes, their use as the only method of recording what happened within a consultation, and the length of time after the event that these interviews took place, suggests that the study design as reported here, is inappropriate. The authors refer to the possibilities that recollections were hazy and that it was likely that other options were entertained but forgotten. They also acknowledge in the limitations section that participants were constructing narratives of events which provided reasons for errors, and point out that the production of stories that make sense, might not match how the decision was actually made, and might not identify what the reasons for errors actually were. However, this weakness in the study design is not adequately addressed. For example, their conclusion that 'lack of knowledge' was an important factor in explaining error, is not convincing, considering their research design. Lack of knowledge would appear a more socially acceptable explanation for error than bias or prejudice for example, and results in a 'better narrative' , but might not accurately reflect the processes through which the error occurred.</p> <p>With respect to the reporting of methods, the article states that "Three researchers searched the text independently for underlying themes in relation to our conceptual framework". However, the analytic method and coding processes were not described in enough detail. For example was every line of transcripts coded? Was constant comparison used? The authors refer to inconsistencies between coders but how were inconsistencies within interviews dealt with analytically? There doesn't appear to be examples of exceptions, or inconsistencies within accounts.</p> <p>However, some of the findings are credible (e.g. the account of cognitive biases developed at framing), and clinicians do appear to provide some less socially acceptable reasons for error - for example, assumptions about a patient's psychosocial history or credibility.</p>
GENERAL COMMENTS	<p>The main messages communicated by this paper are important ones. The 'summary of how warning signs may be incorporated in a practice environment' in figure one might also be of value in clinical practice. However, the main messages, and something very similar to the summary provided in figure 1, could probably have been made on the basis of the current literature. Whereas some of the findings are credible, others are less so. Future studies, adopting some of the revisions to design suggested by the authors, would</p>

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. The checklist states that saturation was reached at about 2/3 of way, this does not appear in the manuscript:

the following has been added to the Methods section under Participants: We agreed that data saturation had been reached about two thirds of the way and did not recruit more participants.

2. It may be appropriate to state that all 15 GPs gave informed consent to use their data:

the following has been added to Appendix 1: Before commencing the interview participants were reassured about confidentiality and advised that tapes would be destroyed so that voices could not be recognised. Further questions about the process were invited. They were also assured that there would be no way that patients could be identified. After Question 1 they were again asked if they were comfortable to proceed.

3. Sometimes difficult reading:

We hope that the changes we made following the Reviewer's suggestions will make for easier reading.

4. Introduction:

4.1. The paper will become more readable if it is seriously shortened:

We deleted the following section from the introduction: The literature suggests that clinicians should reflect on their performance to identify places where their reasoning may be most at risk of failing. This needs to be based on how we actually think rather than how so called rational models work. We need to organise the relevant principles of cognitive psychology in ways that are useful to alert clinicians to possible cognitive bias and emphasise during the training process. These alerts to possible biases may then guide the clinical reasoning process to reflection.

4.2. The sequence of the terms differs in line 1-3 from line 4:

We deleted line 4: Now, generally referred to misdiagnosis, missed diagnosis, or delayed diagnosis.

5. Methods:

5.1. Data saturation is not mentioned in the paper (the information is available in the checklist): see 1 above.

5.2. § 2: this paragraph is difficult reading. It would be possible to reformulate this in more simple and direct language: Replacement of paragraph 2: Closure of the process may occur with minimal or no exposure to the analytic System 2. Just how much analysis occurs depends on the relative dominance of System 1. The use of closure rules, [19] is consistent with System 2. Diagnostic error frequently occurs at initiation, but most often at closure and both Systems 1 and 2 may be involved. [4] It is the congruence of the DTC with what we know about clinical reasoning that leads us to use it as the model for this study.

6. Results:

6.1. Line 3-5 from the end (Do I want ... mayhem): I don't understand this sentence:

corrections appear in red: contributing to faulty decisions: [knowing how busy it is before a week end]

do I want to send a frail, elderly lady up to the hospital on a Friday afternoon when it would be mayhem [there](34);

7. Discussion: did not find a figure:

Figure 1 is at the end of the main document just before the check list.

8. Conclusion: Do you really consider this section necessary? I would suggest to drop it. Nothing new is presented:

in our experience a lot of people read the Conclusion first and then decide whether to go further. We agree there is no new material there, but some people like a summary at the end, so we would prefer to leave it as it stands.

9. Tables:

9.1. Table 1 / second column: your message would be clearer if you add to each sentence something like "instead of ..." with the correct diagnosis:

We hope the Reviewer would accept reasoning, but we would prefer not to do this. Our reason is that we are worried that readers would get stuck on the rights and wrongs of a specific case. Rather, we would hope they would think of the non-specific quotes in relation to their personal experience.

9.2. Box 2: Reconstruction / Presentation, 1th §: What is PR? When were PR and bloods performed? Six months ago or now? Analysis / Presentation / second §: Are you sure you didn't switch around rule in and out?

1th §: changes appear in red: no rectal masses on rectal examination. [Also] did some bloods
second §: correction in red: normal Hb wrongly used for rule out

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. It is well recognised that retrospective interviews are an unreliable method for capturing what actually went on within an interaction or institutional practice. Whereas interviews might have an important role in exploring cognitive processes, their use as the only method of recording what happened within a consultation, and the length of time after the event that these interviews took place, suggests that the study design as reported here, is inappropriate:

We accept the limitations of studies where data collection is limited to interviews. We stress this throughout the report, particularly in the Strengths and limitations section. To make this more forceful, we inserted the following at the beginning of the METHOD:

As indicated in the introduction, the methodology of the majority of studies of diagnostic error rely on retrospective reviews of data collected for malpractice claims. A lesser number of reports deal with hospital incidents and consist of ill defined cohorts. In these reports a mix of poorly defined methodologies are used. As to what happened during the actual consultation is rarely understood and by the nature of the data, unavailable. Instead, chart reviewers tend to make personal judgments. A further confounding problem relates to the unavoidable fact that clinicians may never find out about their errors and if they do, this could be weeks or months after the event. These issues will be discussed in more detail in the section on Strengths and limitations of the study reported here. Our choice of method, relying on interviews for data collection was made with the aim of getting a little closer to the truth of what happens when a consultation effectively fails. Knowing the limitations of the method, our expectation was that this may lead us and hopefully others, to further fruitful research.

We leave discussion of these matters to the section on Implications for further research, at the end of this paper.

2. The authors refer to the possibilities that recollections were hazy and that it was likely that other options were entertained but forgotten. They also acknowledge in the limitations section that participants were constructing narratives of events which provided reasons for errors, and point out that the production of stories that make sense, might not match how the decision was actually made, and might not identify what the reasons for errors actually were. However, this weakness in the study design is not adequately addressed:
see above.

3. For example, their conclusion that 'lack of knowledge' was an important factor in explaining error, is not convincing, considering their research design. Lack of knowledge would appear a more socially acceptable explanation for error than bias or prejudice for example, and results in a 'better narrative', but might not accurately reflect the processes through which the error occurred:

We agree that our view may be incorrect and say so in the text. Our view was based on a number of subjects referring to lack of experience causing the error. Knowledge often comes with experience, but of course there are many other implications. But, in clinical work, especially for reflective practitioners like our cohort, knowledge does come this way.

4. With respect to the reporting of methods: the article states that "Three researchers searched the text independently for underlying themes in relation to our conceptual framework". However, the analytic method and coding processes were not described in enough detail:

We changed the Analysis section:

Three researchers searched the text independently for underlying themes in relation to our conceptual framework. The steps in the dual theory provided the structure for this, so that it commenced with themes for the initial salient features of the case and ended with closure. Categories were coded according to emerging themes and added to or changed as new concepts emerged. [20] The researchers compared their findings in person, on Skype and through emails. This occurred up to once or twice a week over a 2-3 month period. Differences in interpretation were few and were discussed. There were no major differences and consensus was reached on clustering to common themes. Direct quotes from the interviews appear in italics, numerics preceded by G are interview identifiers.

4.1. For example was every line of transcripts coded?:No.

4.2 Was constant comparison used?: No.

4.3.The authors refer to inconsistencies between coders but how were inconsistencies within interviews dealt with analytically? There doesn't appear to be examples of exceptions, or inconsistencies within accounts.:
see above.

5. However, some of the findings are credible (e.g. the account of cognitive biases developed at framing), and clinicians do appear to provide some less socially acceptable reasons for error - for example, assumptions about a patient's psychosocial history or credibility.

The main messages communicated by this paper are important ones. The 'summary of how warning signs may be incorporated in a practice environment' in figure one might also be of value in clinical practice. However, the main messages, and something very similar to the summary provided in figure 1, could probably have been made on the basis of the current literature:

The similarity of many of our findings to reports in the literature was encouraging and suggested to us that eventually they will have more generalisability when more evidence will become available. As for Figure 1, it is possible that it could have been derived from the existing literature, but we have not seen it described. However, we believe that the insights from our study as to what may be occurring in a consultation that leads to error, may have been helpful.

Whereas some of the findings are credible, others are less so. Future studies, adopting some of the revisions to design suggested by the authors, would provide more credible findings.: we agree. This was our hope and are pleased to have the backing of the Reviewer.

VERSION 2 – REVIEW

REVIEWER	<p>Prof Frank Buntinx Dept of General Practice University of Leuven (Belgium) & Maastricht (The Netherlands)</p> <p>I collaborate and published together with one of the co-authors (MT), but on a different topic (ill children) and using different methods (quantitative studies & systematic reviews)</p>
REVIEW RETURNED	16-Jul-2012

GENERAL COMMENTS	<p>The authors partly complied with the (minor) remarks. I can live with the final version. As far as I am concerned, review of a revised version is not necessary anymore</p>
-------------------------	---

REVIEWER	<p>Dr Lucy Brindle Improving Earlier Diagnosis Research Group Lead, Faculty of Health Sciences University of Southampton</p> <p>There are no any conflicts of interest in the provision of this peer review.</p>
REVIEW RETURNED	07-Aug-2012

RESULTS & CONCLUSIONS	<p>The section added to the methods on page 3 improves the case made for the value of the evidence provided by this study. However, a couple of sentences of this section might be rewritten for clarity.</p> <p>On page 4 where you state "It is the congruence of the DTC with what we know about clinical reasoning..." With what you know on what basis? Current evidence? Perhaps you could specify this. This lack of clarity is probably a consequence of writing in the active rather than passive voice.</p> <p>The conclusions you draw about the implication of 'knowledge gaps' for diagnostic reasoning are appropriately tentative in most places.</p>
----------------------------------	--

	<p>However, on page 10 you conclude that "we therefore conclude that lack of knowledge is indeed likely to be an important factor in diagnostic error". Considering weaknesses of your method raised by yourselves and in my previous review I think that this is a problematic conclusion. References to knowledge gaps would provide a relatively neutral way of retrospectively making sense of diagnostic errors. However, your conclusion is in keeping with your broader methodology so it should be the decisions of the journal, and yourselves, whether or not you make this recommended revision. Overall, the article makes a valuable contribution to the literature.</p>
--	---