Author's response to reviews

Title: Predictive ability of an early diagnostic guess in patients presenting chest pain

Authors:

François Verdon (Francois.verdon@mcnet.ch)
Michel Junod (junod.nicod@bluewin.ch)
Lilli Herzig (lilli.herzig@hin.ch)
Paul Vaucher (paul.vaucher@hospvd.ch)
Bernard Burnand (bernard.burnand@chuv.ch)
Thomas Bischoff (thomas.bischoff@hospvd.ch)
Alain Pécoud (alain.pecoud@hospvd.ch)
Bernard Favrat (bernard.favrat@hospvd.ch)

Version: 3 Date: 29 October 2009

Author's response to reviews: see over
Dear Editor, dear reviewers,

Please find our corrections and comments for the manuscript entitled: “Early diagnostic impression has good predictive ability in primary care patients presenting chest pain” for which we have received your suggestions for revision.

We thank you for your critical appraisal of our short report. As many comments have been addressed, we have decided to change the format of the article to a more classic full original version. This makes it possible for both reviewers and future readers to assess the reliability of our conclusions.

Below we provide a listing of each of the comments that you have sent to us in bold, followed by our responses and description of changes made to the paper.

Reviewer: Rudi Bruyninckx

Thank you for the time you must have spent to review our manuscript and for your important contribution in improving our paper.

Introduction:

1. ‘What were the results of the study? Probably it was ‘the positive predictive value of the early diagnosis impression was of 51% (95%CI: 49.4-52.5)’ (Abstract, measurements and main results). If my interpretation is correct, you have to repeat this here and not only in the abstract.

The introduction was revised. We are not sure to have understood your comment. We thought it was indeed important to give more details in the last paragraph on the objectives and specify that we wanted to measure the positive predictive value.

2a ‘Furthermore, we examined if this predictive ability….age and gender’. This is what you did, but this is not a research question. It must be clear what you want to do in this new (part) of the study.

“Is the predictive values similar across different sub-populations?” was our secondary question. We formulated our aims as objectives and not as research questions. To make it easier for readers to find our objectives, we transposed the last paragraph from the introduction to the method section and named it “objectives”.
2.b Please define a research question.

Objectives (primary and secondary) were reformulated in the method section.

Methods

3. Results belong to the result section.

The first sentence was moved to the beginning of the results section.

4. How were the practices selected?

Practices were those from a research network in western Switzerland. They are therefore not a random sample of physicians. Their ability could be different from other practitioners. We therefore added this detail in the method section and added this limitation in the discussion.

5.a Is the ‘diagnostic impression’ the same as ‘the initial, early diagnostic impression that includes initial contact, the first words of complaints, and patient’s presentation’ (Introduction)? If so, you have to give this definition in this method section and skip it in the introduction.

This was moved from the background section to the method section were we defined what was meant by “early diagnostic impression”.

5.b Which diagnostic impressions were possible?

Physicians were free to report diagnosis as they wanted to. Diagnosis were then categorised in larger categories of diagnosis for comparison with initial impression. These are reported in table 1.

5.c How did you make a difference between: ‘chest wall syndrome’, ‘anxiety disorder’ and ‘somatisation’? Maybe this is something for the discussion?

Physicians were free to diagnose the pain’s origin as they would in usual care. A panel of experts then examined the diagnosis retained at the end of the first encounter, at three months and at one year. When incoherence appeared, the panel would use the patient’s file and discuss the most probable diagnosis. It is however indeed sometimes difficult to distinguish chest wall syndrome from psychogenic pain; this was added to the discussion.

6. ‘the expressed patient’s anxiety within the first two minutes’ Could you give some more information on how the GPs have to examine this? Were there rules or was it just the GP’s impression?

To assess the patient’s anxiety about their chest pain, the physician had to report “yes” to the following statement: “The patient explicitly expressed worry about his pain.” This was added to the method section.

7. ‘The precise diagnosis’ Probably the same disease categories are used as for the ‘initial diagnostic impression’. I suggest to mention that.

Indeed, the same categories were chosen for the “final diagnosis”. This was added to the method section.

8. ‘if any’ Please add which ‘definite diagnosis’ was used when there was no further information during the follow-up. This is also a point for the discussion section because the initial and the final diagnosis is made by the same GP.

The final diagnosis was validated by a panel of GPs who used the information reported by the physician without using his initial diagnostic impression. For the few patients that were to be lost at follow-up, the information collected during the study and the patient’s up to date file were used for the panel of experts to define the definite diagnosis. 100% of patients were followed-up at 3 months and 96% at one year. Using delayed diagnosis decreases the chances of having falsely ruled out an important diagnosis. However, the lack of independence
between the initial diagnosis impression and the information used to define the final diagnosis introduces a potential bias. This is stated in the limitations section of the discussion.

9. ‘Predictive values of first glance impression’. I am not sure that you may use ‘predictive value’ here. For me they both exist i.e. a positive and a negative predictive value, being the results of the prevalence of a disease (like anxiety disorder) and the accuracy (sensitivity and specificity) of a test (here: first glance impression) using Bayes theorem. So I prefer ‘predictive ability’ over ‘predictive value’. I think you did so: a GP made 100 initial diagnoses of anxiety disorder and after one year the precise diagnosis of those 100 patients was confirmed in 45 patients, and so on for all disease categories. But also possible: a GP made 100 final diagnoses of anxiety disorder and then compared these with the initial diagnosis. After all, how you get the result is not clear for me and I need more information.

The term “predictive values” was changed to “predictive abilities” as suggested. The method used was delayed diagnosis. This means we confirmed the diagnosis that was suggested at the end of the first encounter by following patients over one year. If new investigations, or if new clinical manifestations suggested a different diagnosis which was then confirmed, the initial diagnosis reported by the GP at the end of the first encounter was considered as inappropriate and was replaced by the new more relevant diagnosis.

10. mean value with 95% CI. You mentioned only once this interval in the result section. Add the 95% CI when useful.

95% CI are given to infer results to what could be found had we repeated the study a number of times. It is therefore important to only report 95% CI for measurements which have been chosen to answer the studies objectives. We have given 95% CIs for the prevalence of correctly predicted diagnosis and for ORs in the stratified analysis. Standard deviation and sample size are given for other measures for use in meta-analysis.

11. ‘Significance level was set at p<0.05’. You announce this here but it is not used in the text, so don’t mention it at all.

Significance level to reject the null hypothesis has to be stated in the statistical method section. p-values were therefore added in table 3 to prevent any confusion.

Results

12. ‘In patients lost during the follow-up at one year, the 3-month diagnosis was used.’ This is no result, this is for the method section.

This sentence was deleted and precisions were given in the method section.

13. ‘Patients with missing data were excluded.’ This is no result, this is for the method section.

No precisions were given in the protocol on how missing data was to be handled. The choice of reporting results only for patients with full data was made once we were certain this did not introduce a bias. This can only be done once the data is available. We therefore believe this statement should be in the result section and not in the methods section.

14. ‘and most physicians were men.’ If you have information on the localisation of the practices (urban, rural) add it. Suggestion: make a table with the GP’s characteristics.

As suggested, we added a table (Table 2) giving more precisions on physicians’ characteristics.

15) ‘...were seen with chest pain (51.4%) and 51.9% of patients were over 50 years.’ Add the 95% CI or make a table with patient characteristics and the 95% CI.

95% CI should not be used for descriptive purposes unless the objective of the study is to describe the population. We therefore disagree to report 95% CI for those characteristics which would be confusing for readers. The numerator is nevertheless known to the reader who, if needed, can calculate the 95% CI.
16. “chest wall syndrome for 266 patients (42.5%), coronary...for 62 patients (9.9%). Add the 95%CI or make a table with patient diagnoses and the 95%CI.

Again, we disagree with the reviewer on the use of reporting 95%CI for variables describing the studied population. We do not have any intention to infer these characteristics to larger populations.

17. ‘An initial diagnostic impression was recorded in 441 patients (70.4%) and confirmed in 319 patients at one year.’ I do not understand this: you had full data of 626 patients and now only 319. You have to explain this.

Not reporting any diagnostic impression was not considered as a missing diagnosis impression. It only meant they did not have any impression to report. Predictive ability for those patients was therefore considered as null. Not including these patients in the analysis would have introduced an important bias in our results. We have reformulated our statement in the result section to make things clearer.

18. “showed little heterogeneity between GPs. Was there a difference between the older and the young GPs (less than 2 year of experience)?

No GPs with less than two years of experience were included in our study. This question therefore remains unanswered. There was a non-significant trend for more experienced physicians (≥10 yrs practice) to show a better ability to correctly predict the final diagnosis (OR=1.5 CI95% 0.89; 2.5). This was added in the result section.

Discussion

19. I prefer a structured discussion: summary, strengths and weaknesses, previous studies, future research.

The discussion was structured as suggested.

20. ‘We consider that thoracic pain in primary practice is...life-treating condition.’ Suggestion: better in the introduction.

This sentence was moved to the introduction which was made more complete.

21. ‘Furthermore, at first glance, GPs were also able to detect two thirds of life treating diagnoses.’ Is this a good result? Yes and No. If you are recognised immediately it is good, if not it is bad. What is the importance of the first glance diagnosis? What is the importance of history taking and the physical examination to make a better diagnosis.

This has been added in the discussion using results from other studies including those you suggested to us.

22. ‘Thus, the initial diagnostic impression may be obtained by means of non-explicit pathways, based on intuition, associations with stored information an pattern recognition. True, but we also got a lot of information at the first glance: is at the GP’s office, is it an urgent home visit, the age of the patient, gender, is patient lying in bed or on the ground, or normal walking, is he pale, is he anxious, are the relatives anxious, .. If it is only stored information and pattern recognition, younger colleagues must have worse results. Is this so?

This what not the case in our observation. We have therefore taken your comment into consideration and have added it to our discussion.

23. ‘To our knowledge, this is the first cohort study to examine the performance of the first diagnostic impression.’ Our group in Leuven, Belgium also worked on chest pain. Of course it is not exactly at the first glance diagnosis, but we suggest to read them. Please also read Abu Hani and my qualitative study of patients with chest pain. My colleague, Ann Van den Bruel made a similar study on the diagnosis of serious infections in children.

Thank you for sharing these references with us. We have studied them and have included most of them in our discussion.
Table 1
24. ‘SE’ Please replace this by the 95%CI.

SE was mistakenly written instead of SD. This was corrected. Table 1 is now called table 3. As this was not our chosen measure of effect, we prefer not to give confidence intervals. Those have been given for ORs. However CI95% were added in the text as suggested.

Abstract

25. “life treating affection (65.4%) and in patients who did not feel anxious (62.9%)’ Please add the 95%CI.

CI95% were added.

26. ‘GPs were able to rule out a majority of life treating diagnoses at first glance…’ You can certainly rule in a lot of life-treating diagnoses, but how much diagnoses are we going to miss by ruling out on the basis of the first glance? So I am not convinced you can state this. You have to discuss this in the discussion section, perhaps together with remark

This sentence was withdrawn from the abstract and the subject was discussed in the discussion section.

Reviewer: Ana Ruigomez

We appoligise for not having made it clearer that our draft was a short report. We have now included more details in this version and hope to answer most of your questions.

Major Compulsory Revisions

1. There is a need for a justification of the study. Why they study the first glance diagnosis just based in the first few minutes of consultation? Why not considering clinical symptoms, presentation and information provide by the patient during consultation, that could be useful to establish an early diagnosis? Why not considered basic physical examination?

The introduction was rewritten and made clearer. Furthermore, the TOPIC study also investigated diagnostic abilities after history taking, after physical examination and after laboratory investigations. This has now been reported in this paper through figure 1.

2. What is the extra value of “a first glance diagnosis” that could be seen as a suspicion or just a guess versus a diagnosis based on the impression after the encounter /consultation has finalised (taking into account the symptoms, and history given by the patient)?

This study is not meant to show that first glance is an appropriate mean to diagnose an affection. Our aim was to show that other factors than those yet studied in evidence medicine is apparently used by GPs to orientate their investigation and seems to play an important role in the management of patients in primary care. We have made this clearer in our introduction.

3. Methods. Need to be re-written.

The entire section was structured and rewritten as suggested.

4. Exposure: Initial suspicion of diagnosis before history taking. Is this a just a guess? Not clear definition and how it is measure.

An entire paragraph was written to describe how the initial diagnostic impression was measured.
5. How are they sure they measure or capture this first impression and not the impression after consultation considering rest of the facts? Did the GP stopped their consultation at a certain timing and filled a questionnaire and then continue with the consultation?

The physicians had to fill in the CRF at different timings of the consultations. They were trained to do so and therefore filled in their diagnostic impression before ongoing further investigations. We therefore cannot exclude that physicians did not behave differently to what they usually do as they were filling in the CRF. This was added to the limitation section of the discussion.

6. How was the procedure to ascertain the initial suspicion of diagnosis? Not enough details are provided to understand how they collected information from the GP.

The method section was rewritten to add all these details.

7. I read in a previous publication from authors (BMC Fam Practice, 2007, 8:51), that GPs had to filled a specific questionnaire, but it is not mentioned or describe in the present manuscript. Also a description of information about diagnosis at follow up it also lacking. Do they use the same method (questionnaire) to assess 3 month and 1 year diagnosis? Was the same questionnaire? Or used medical record extraction?

Methods used to define the final diagnostic was described in details in a subsection from the method section.

8. Not clear the number of patients that were used for final analysis. “Initial diagnosis impression was recorded in 70.4% of patients and confirmed in 319 patients at one year”. This is hard to understand. What happened with the rest?

Physicians were not obliged to give a diagnostic impression. They did so only if they had a presumption. This was confusing. The section was written over again to make this clearer. (also see comment 17 for reviewer 1).

9. Missing information? Were they excluded from the analysis?

Denominators have to be clear for readers. We either could include different sample size for each studied variable or decide to only keep full data. This was done to make things easier for the reader. We however verified that conclusions remained the same using full data.

10. Statistical analysis. Unsure about statistical methods used. They do not describe how they construct the logistic model, unique or different models? Which variables of adjustment were included?

We did not construct a logistic model. Random effect logistic regression was used to adjust for lack of independence related to the fact our outcome was associated to each physician. This is one of the best statistical methods used to make it possible to adjust the significance level and the magnitude of effects for clustering effects. The only variable the results were therefore adjusted for is the physician level. In other words, we corrected the lack of independence of the measures which makes it impossible to use standard methods of inference.

Minor Essential Revisions

11. Table 1 (SE) means standard error? Why in %?

Thank you for spotting this mistake. It should have been stated SD and not SE. We have calculated the predictive ability at a physician level. This means we aggregated the data at a physician level separating observations between those with or without the exposure. We then calculated the mean value over all physicians. This means that the inference is over a quantitative variable which corresponds to the predictive ability of each physician for each sub-populations.
We hope our comments and modifications have satisfied you. Thank you again for your contribution in improving our paper.

Sincerely,

Paul Vaucher