Author's response to reviews

Title: Diagnostic accuracy of point-of-care testing for acute coronary syndromes, heart failure and thromboembolic events in primary care: a cluster-randomised controlled trial

Authors:

Yuki Tomonaga (yukit@access.uzh.ch)
Felix Gutzwiller (gutzwill@ifspm.uzh.ch)
Thomas F Lüscher (karlue@usz.uzh.ch)
Walter F Riesen (Walter.Riesen@ikch.ch)
Markus Hug (markus.hug@hin.ch)
Albert Diemand (a-l.diemand@bluewin.ch)
Matthias Schwenkglenks (m.schwenkglenks@unibas.ch)
Thomas D Szucs (thomas.szucs@unibas.ch)

Version: 2 Date: 25 June 2010

Author's response to reviews: see over
MS: 6806930873847505 - Clinical benefit of point-of-care testing for acute coronary syndromes, heart failure and thromboembolic events in primary care: a cluster-randomised controlled trial.

Dear Mrs Pafitis,

We have specified the name of the ethics committee which approved the study and we have revised our manuscript according to the reviewer comments. Please find on the next pages our point-by point responses.

Sincerely

Yuki Tomonaga, MSc.

Prof. Dr. med. Thomas D. Szucs

25.06.2010, Zürich - Switzerland
Point-by-point response
Reviewer: James Nichols

1) The study should have included traditional biomarkers and other tests from a clinical laboratory (even though remote from the clinic) as part of the working diagnosis rather than just history, symptoms and physical findings. By providing POCT in the working diagnosis, the study wasn’t necessarily comparing speed of test availability (i.e., biomarkers in the clinic versus in a distant lab) but rather clinical judgment with and without biomarkers. As such, these biomarkers have already been well studied and characterized with respect to diagnostic accuracy. This point should be made in the discussion as one potential source of bias in the study design.

Reply: We agree that the main comparison was clinical judgement with and without biomarkers. Of course we expected to see a difference between the groups, since it is well known, how accurate such tests are (we also cited several papers that confirmed this accuracy). What is not really known is how much these biomarkers can contribute in a general practice setting (in the majority of available studies, only one biomarker was investigated in a highly selected group of patients). With this study we are able to give a first estimate of the real effects of POCT devices in primary care. Therefore there is no bias in the study design.

2) Another issue that should be further examined is the panel of tests versus availability of single tests. The POCT provided to the clinics allowed analysis of 3 biomarkers at the same time, troponin, BNP, and D-Dimer, yet the authors chose to only statistically evaluate the diagnostic efficiency of each marker separately for cardiac syndromes, CHF, and thromboembolic events. What, then, was the value of performing all 3 tests as a panel on all patients? If the doctor suspected ACS, then troponin was really only necessary. This fits with the US insurance payer system where panels of tests are not reimbursed and each single test must be justified medically for reimbursement of the test cost. Since the authors performed all 3 tests on all patients, then the authors should have evaluated whether the knowledge of the other two tests increased the diagnostic accuracy of the primary test for the working diagnosis? In other words, if troponin would have predicted ACS with some degree of efficiency, did knowing that BNP or D-Dimer were positive add additional diagnostic accuracy to the working diagnosis? This should be evaluated and discussed.
Reply: This is a very good comment. And to be honest, by reading again the manuscript, it seems that all biomarkers were measured for every POCT-patient. However, this was not the case. (In the Swiss insurance system it is also necessary to justify the analyses performed on a one-by-one basis ...) As written at the end of the introduction, “We hypothesised that POCT testing for cTnT, NT-proBNP, and D-dimer...” The POC-device can measure all biomarkers, but not at the same time: for each marker analysis 8-12 minutes are necessary. POCT-physicians had the possibility to analyse all biomarkers, but it was not mandatory. As remarked by James Nichols, it is not necessary to perform a d-dimer or bnp test if the troponin is positive: the risk of myocardial infarction is high and the patient will be hospitalised as soon as possible. In our study, the physicians were free to decide if and which tests were necessary. Given Table 5 it is possible to calculate that 147 (26+121: 67%) patients received a cTnT test, 70 (35+35: 32%) patients a Nt-proBNP test, and 118 (20+98: 54%) patients a D-dimer test. All tests were only performed for few patients. In order to avoid similar misunderstandings in the future, we have now specified in the methods section that the physicians had the possibility to choose if and which biomarker test was necessary for the patient, depending on physical findings, symptoms and patient history.
Point-by-point response
Reviewer: Caroline Laurence

- **Major Compulsory Revisions**

  A major problem with this paper is the inadequate description of the methods. This in turns makes it difficult to assess the validity of conclusions drawn and the results. Specifically the methods do not describe adequately:

  - What is the conventional diagnosis employing best clinical practice? It was not clear what the process was for the control group and a brief description of what this is would help.

  Reply: To diagnose employing best clinical practice just means to diagnose the patients using nationally or locally agreed guidance and available evidence on the clinical provision of healthcare (i.e. the standard, conventional diagnostic process). We think it is not necessary to explain how physicians usually work (or how they should usually work).

  - Why were practices only included if they were some distance from a laboratory with specialised diagnostic systems?

  Reply: We expected to find a more clear cut difference between the study arms. We assumed that control practices situated close to a diagnostic laboratory (which are often in a hospital) would often send a courier with a blood sample to the laboratory, wait for the test results, and then make the working diagnosis. In case of worsening of the patient, the hospital would immediately be reached. Such an effect would totally mask the effects of POC testing.

  In contrast, in practices “far away” from hospitals (which are frequent in mountainous Switzerland), decision-making at the physician's office may have more drastic consequences. In the absence of a POCT device, a patient with chest pain and a suspected myocardial infarction may be sent to a hospital immediately, since it could be dangerous to wait for an external laboratory test (which will probably take a lot more time if compared to practices situated close to a hospital).
• When PoCT was undertaken in the practice? For the intervention group was this done at the first visit of the patient and did the control group have a laboratory test undertaken. Who undertook the test? Was venous blood used? Was there standardisation in procedures between practices? (training of staff in devices, QA program etc)

Reply: As described in the methods, POCT was undertaken before making a working diagnosis. Control group patients did not have a laboratory test undertaken.

We have specified in the discussion (ACS diagnosis): “In our study, training on the POC-instrument always was performed by the same specialist from Roche Diagnostic, but some staff members were instructed by their colleagues only.

As written at the end of the introduction: “We hypothesised that POCT testing for cTnT, NT-proBNP, and/or D-dimer in venous whole blood ...

As written in the discussion (ACS diagnosis): “Test quality was monitored using the internal and external quality controls required by Swiss federal law and the Swiss Commission for Quality Assurance in the Medical Laboratory (QUALAB)”

• Where the intervention GPs provided with information on the interpretation of the results?

Reply: All GPs who received a POCT device received advice on the interpretation of test results. We added this information in the methods.

• It is not clear how the confirmed diagnosis was reached. Who determined the confirmed diagnosis? Was it an independent physician or local hospital? What evidence did they use for this? Is the confirmed diagnosis then used to determine if the working diagnosis was correct?

Reply: As written in the methods: “At follow-up 3 weeks later the same physician reviewed the working diagnosis. The follow-up diagnosis was defined as the confirmed diagnosis. Follow-up data of patients requiring additional specialist visits or hospitalisation were provided by specialist or hospital reports.”

This is of course a limitation of this study, as we explained in the discussion “The follow-up diagnoses were another limitation. In principle, they should have been performed by independent blinded assessors but this was not feasible. For practical and data protection reasons, they were made by the same physician who was primarily consulted by the patient. We are aware of the potential for bias, e.g. due to possible underreporting of incorrect baseline diagnoses leading to a false high rate of correct baseline diagnoses or a false
low difference between study arms. On the other hand, for patients referred for further diagnostic work-up (including all patients at potentially high cardiovascular risk), GPs received a written report on the second-stage assessment as well as information on further clinical management, which substantially reduced the risk of bias.”

- The baseline data collected from the patient is not mentioned in the methods, but reported in the results.

Reply: We specified that the GP made a working diagnosis of ACS, HF, TE, musculoskeletal or “other” (specified) problems based on the patients characteristics, history, etc.

The aim of the study did not match the methods and result. The hypothesis was that PoCT testing for cTnT, NT-proBNP and/or D-dimer in venous whole blood would provide simple, rapid and accurate diagnosis of ACS, HF and TE. Nowhere does the paper report on the how the process is simple or rapid. It is only dealing with the accurate diagnosis. The hypothesis needs to be refined to reflect what is reported. Additionally, it should include what its accuracy is compared to or is the aim improving the accuracy of GP diagnosis through the use of PoC tests? Is accuracy being defined as the correct diagnosis by GP (and confirmed) as well as accuracy of the specific tests in identifying the condition? Or both?

Reply: We corrected the title and aim: in fact, in this manuscript, we compare the diagnostic accuracy between GPs using POCT and controls: “We hypothesised that POCT testing for cTnT, NT-proBNP, and/or D-dimer in venous whole blood would allow for a more accurate diagnosis of ACS, HF and TE by office-based, Swiss general practitioners.”

The “simplicity and rapidity” are a direct consequence of the fact that the POCT device provides fast analyses and is really simple to utilise. However we agree with the reviewer: using the terms “simple and rapid” in the hypothesis could mislead the lectors. Therefore we have deleted them.

In the paper the diagnostic accuracy is defined as correct diagnosis by GP. Additionally, in table 5 it is possible to see how accurate the specific tests were. As already mentioned, the biomarkers were just an additional tool used to make the working diagnosis.

Some interpretations of the results are inaccurate and misleading and need to be revised. In results, para 1 the authors state there was a significantly higher LDL value in the controls. There is no evidence of a statistical test being undertaken and
so the term significance needs to be removed. Moreover, if the LDL is an example where the intervention and control differ, why is not glucose included as the difference is the same as for LDL (0.4 mmol/L) ?

Reply: In the baseline characteristics it is often not necessary to include a column with the significance levels. However, since one parameter was significantly different between the groups, we decided to cite this difference in the text. In order to avoid similar feedbacks in the future, we have deleted the comment on higher LDL.

Concerning the “glucose”-comment, we would like to remark that the significance of a difference of 0.4 depends on the calculated means (and on the number of included subjects). In our case, a LDL increase of 0.4 mmol/l indicates an increase of 13.7% (P=0.018). For glucose, an increase of 0.4 mmol/l represents just 6.7% (P=0.142).

In the results para 3, the authors state that the groups differed in rates of acute chest pain and calf pain. For calf pain the difference between the groups was 4%. If this determines a difference then why was not Oedema also included with a 4% difference? Similarly in Table 3/results para 4, the authors state that only the proportion of patients with previously diagnosed HF differed substantially between groups. In reviewing the results in Table 3, a greater difference was found between the treatment groups for hypertension (7% difference) and diabetes was 6% the same as HF.

Reply: Statistical analyses has shown that the groups differed in rates of acute chest pain and calf pain. The 4% difference for calf pain (7 vs 3%; P=0.039) has not the same level of significance than the 4% difference for oedema (10% vs 9%; P=0.111).

To make sure that the readers will understand that the mentioned differences are based on statistical tests (which is an obvious consideration), we specified that “The groups statistically differed in rates of acute chest and calf pain.”

Table 4: I am not clear about what is presented with this table. Is it showing the accuracy of the working diagnosis based on the confirmed diagnosis for each the treatment groups? If it is, this needs to be made clear in the heading and the text.

Reply: Yes it is. We changed the heading and specified that the table shows the accuracy of the working diagnosis based on the confirmed diagnosis for each of the treatment groups.
I am not sure if the conclusion regarding the greater diagnostic accuracy in the PoCT group (results para 8) are correct. For three of the conditions the NPV was almost the same 100% vs 99%. It was only with the MS and other conditions where there seemed to be some difference between GPs using PoCT for their diagnosis and those who did not. However, this small difference may be clinically important and so needs to be interpreted in this context.

Reply: It is true that the negative predictive values were almost identical between the groups. In contrast, sensitivity and specificity were different and generally higher in the POCT group.

As mentioned in the text, 76% of the POCT diagnoses were correct (compared to 60% in the controls). Working diagnoses of ACS, HF and TE proved correct in 70% vs 45%.

This clearly indicates a greater diagnostic accuracy in the POCT group.

We think it is not necessary to discuss the differences in MS and other conditions: these are simply due to the fact that in the control group there were more false positive diagnoses of ACS, HF, and TE. We also wrote that “correctness of the remaining working diagnoses (musculoskeletal or “other” problems) did not differ: 80% vs 73% (p=0.31)“.

Some clarity about the biomarker performance is needed (results para 9). This only relates to the intervention group of results and if so, this needs to be stated in the text. I am unclear as to why the working diagnosis and confirmed diagnoses are included. If the aim is to test the accuracy of the three tests to identify correctly the diagnosis then it should be compared to the confirmed (correct) diagnosis. The inclusion of the working diagnosis results seem to indicate that the GP may not have been interpreting the test results correctly when making his/her working diagnosis and so it is not an issue of the accuracy of the test, but the use of the test in making a clinical decision. This is a different question entirely and an important one to discuss.

Reply: Although we would regard it as common sense that the performance of the biomarker tests is exclusively assessed within the intervention group, it is better to state it in the text. We changed the heading in “Biomarker performance in the POCT group”.

The remark about the presence of both confirmed and working diagnosis is very helpful. To be honest, we had planned to address this aspect in the discussion, but skipped it due to word count limitations, in the initial submission.

Like Dr. Caroline Lawrence said, the test accuracy should be just compared to the confirmed diagnosis. We decided to include also a comparison with the working diagnosis since we thought that it is an interesting remark. Of
course we can and will not judge the GPs work, but it seems that in some cases the GP decided to ignore the results of the biomarker tests. Maybe they were not totally confident with the results, and/or they decided to give more importance to the physical findings, symptoms, and patient history. It would be interesting to make a study to investigate how much the GP trust the biomarkers. We added a paragraph about this topic in the discussion.

Discussion – overall the discussion is long and some of the conclusions drawn are overstated based on the result provided. This is made more difficult by the lack of information provided in the methods, which makes interpretation of the results very difficult.

Reply: We agree that the discussion is long. However, in the majority of comparable studies only one marker/disease is investigated… in our manuscript we have three of them. Each needs some separate discussion with comments and comparison with the literature. We changed some adjectives in order not to overstate the conclusions.

The conclusion need reworking as the study did not report on the simplicity or speed of the PoCT devices an only looked at the role in accurately determining the diagnosis.

Reply: We changed the conclusion: “…analysis of cTnT, NT-proBNP and D-dimer produced more accurate diagnoses of ACS, HF and TE in the POCT group”

Minor Essential Revisions

Table 3 title and headings need to be changes as all items listed are not pre-existing diagnoses.

Reply: Title and heading changed in “medical history”

If the focus of the paper is a comparison between groups (intervention and control) it is not clear why in the results presents the frequency of diagnoses within each group (results para 5) While it is of interest, it does not contribute to the paper.

Reply: We believe that it is important to show the frequency of the diagnoses within each group. Figure 1 shows that by the working diagnoses the diagnostic frequencies in the two groups were significantly different. How-
ever, the confirmed diagnoses showed no significant difference between the groups. This suggests that there was no selection bias in the study groups. We added a paragraph on this topic in the discussion (limitations).

Results para 7. It is not clear in this paragraph which is the intervention and control group results. It would help if this was included in the text and also the frequencies included.

Reply: we have specified which are the POCT and which are the control frequencies.

Not clear what ** means in presentation of p values.

Reply: In many scientific publications, the ** (asterisks) are employed as a shorthand to denote the statistical significance of results. Popular significance levels are P<0.05 (*), P<0.01 (**), and P<0.001 (***)

Since we also wrote the exact P value, the asterisks are not really necessary. Therefore, to avoid potential misunderstandings, we have eliminated the asterisks.

Discussion (para 2) – incorrect result presented in last sentence it should be 78% of controls not 80% of controls.

Reply: right. We thank the reviewer for the careful correction.

Figures 1 and 2. These graphs may be easier to interpret if the intervention and control results are adjacent to each other, rather than in two separate graphs.

Reply: Not really. Since the number of patients included in the two groups was different, it is not preferable to put both intervention and control results in the same graphic. In the actual representation it is not only possible to have an overview on the number of diagnoses in each group, but it is also possible to see how frequent a diagnose was made. Thank to the frequency distribution it is possible to directly compare the two groups.