PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The value of biomarkers in predicting mortality in older medical emergency department patients: a Dutch prospective study
AUTHORS	Zelis, Noortje; Hundscheid, Robin; Buijs, Jacqueline; De Leeuw, Peter; Raijmakers, Maarten; van Kuijk, Sander; Stassen, Patricia

VERSION 1 – REVIEW

REVIEWER	Narani Sivayoham Emergency Department St George's University Hospitals NHS FT London UK
REVIEW RETURNED	23-Aug-2020

GENERAL COMMENTS	Thank you for asking me to review this paper on 'The value of biomarkers in predicting mortality n older medical emergency department patients'. The subject matter is clearly a useful one. However, I have a few comments regarding this study: Major concerns: 1) There does not appear to be a sample size calculation 2) The authors have stated that the venous blood samples were spun down and saved for analysis several weeks or months later. Could the authors confirm if this is standard procedure and that this methodology does not affect the quality of the results? For example, was there a difference between the results measured using this procedure compared to the results obtained when the clinician treating the patient requested it in the clinical setting? 3) In this study the authors have treated each biomarker independently and found that each had a 'sufficient' or 'good' AUC to predict mortality. Have the authors considered using the biomarkers as a panel and studied their collective predictive accuracy? If not, why not? 4) If the collective panel of biomarkers is predictive of mortality, how does it compare with the new simplified model [Ref 17] they have created? Do any of the biomarkers either individually or collectively add value to their model in Ref 17? 5) It is concerning that over 15% of patients were missing a lactate, when only 5-10% of other biomarkers were missing? Why is this? 6) The conclusion the authors have arrived at is correct based on the analysis they have carried out but I do feel the analysis is inadequate for reasons given in comments 3 & 4. Minor concerns: 1) In the introduction section, use the term emergency department once before it is abbreviated to ED.

	2) In the methods section, the authors have stated that p-values ≤0.05 were considered significant. Is this a typographical error? It is usually p<0.05 that is considered significant. 3) The flow chart for patient selection should be part of the main paper, not a supplementary file.

REVIEWER	Camille Schwab INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F75012, Paris, France ; Université Paris-Sud, Faculté de Pharmacie, Département de Pharmacie Clinique, F-92296, Châtenay-Malabry, France ; GHU APHP.Sorbonne Université, Hôpital Saint Antoine, Service Pharmacie, F75012, Paris, France
REVIEW RETURNED	27-Aug-2020

GENERAL COMMENTS	Thank you for the opportunity to review this paper. I have some remarks and questions: In the introduction section: Background The authors wrote that the biomarkers are used to "estimate the severity of specific diseases". But these biomarkers are first used to diagnose the diseases, and then the severity. Importance Can the authors better explain why were they interested in biomarkers that are mostly present in acute diseases (heart failure, myocardial injury) and which have fast kinetics, for long term outcomes (30-day mortality or 30-day unplanned readmission)? In the method section: Study design, setting and selection of participants: 1) As this study is part of the RISE UP study, why did it take place only in the Zuyderlann MD and not in all the centers of the RISE UP study? I think the authors should explain why. 2) Why did the patients were included only if they were treated by an internist or a gastroenterologist? I think it conducts to a classification bias as gastroenterologist's patients have very specific diseases. Measurements 3) Can the authors explain why did the samples were analyzed within 4 weeks or 3-4 months after collection? 4) As the samples were analyzed 1 to 4 months after the ED visit, when did the data were collected from the EMR? Did the Charlson Index was calculated after the ED visit, and thus include the possible new comorbidity? Outcome 5) Can the authors explain how they measured the outcome: - How were the deaths verified? - Are the ICU/MCU admission had to be in the Zuyderlann MD hospital, or in any another hospital, right after the ED visit or within 30 days after the ED visit? - For the LOS, was it an admission in the Zuyderlann MD hospital, or in any another hospital, right after the ED visit or within 30 days after the ED visit? - How was measured the loss of independent living? - How was measured unplanned readmission? Was it unplanned readmission in Zuyderlann MD hospital, or in any another hospital?

	Analysis 6) “Logisitic regression analyses were performed on data after imputation of missing values”. The authors have not conducted sensitivity analyses to make sure of the validity of the imputation. 7) Contrary to what I read in the STROBE statements, I have not read in the manuscript how the study size was arrived at. 8) “Linearity was visually checked”, could the graphs be presented in the supplementary data? In the results section: Characteristics of study subjects 1) “During the follow up period”, as I wrote earlier, can the authors explain in the method section how did they follow the patients? 2) I think the chart inclusion is not complete. Can the authors add the outcomes in this flow chart? 3) In the table 1, the authors have not mentioned the missing values. Can they add the percentage of missing values for each variable so that the reader can better understand results? 4) “Non-survivors more frequently experienced the composite endpoint (72.5%)…”: Can the authors had the N with the percentage, and had this interesting result in the table 1? 5) In the table 1, we can see that the survivors are very different from the non-survivors (age, community-dwelling, Charlson index), but the authors have not taken these differences into account for the logistic regression analyses, and presented non adjusted OR. Main results 6) “The highest values of the biomarkers were more often….between the non-survivors and survivors”: Have the authors statistically measured this result? 7) Have the authors evaluated the predictive validity of the biomarkers of each outcome of the composite outcome, individually? In the discussion section: 1) “We detected no evidence for selection bias but cannot exclude it either” How did the authors evaluated the absence of selection bias?

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Narani Sivayoham
Institution and Country:
Emergency Department
St George's University Hospitals NHS FT
London
UK
Competing interests: None declared

Please leave your comments for the authors below
Thank you for asking me to review this paper on 'The value of biomarkers in predicting mortality in older medical emergency department patients'. The subject matter is clearly a useful one. However, I have a few comments regarding this study:

Thank you for your thorough review. Your comments were constructive and helpful for improving our manuscript. Please find our responses to your remarks below.

Major concerns:
1) There does not appear to be a sample size calculation

The sample size of this study depended on data from a previous prospective cohort study, the RISE UP study, which provided the current study with data of 450 patients. For logistic regression analysis, at least 10 event per candidate predictor are needed, according to prediction modelling guidelines. We assumed that the mortality rate was around 11% and therefore, inclusion of 450 patients would be more than sufficient. In the previous version of the manuscript, we did not elaborate on the sample size. In the revised version, we added the following to the analysis section (p 8):

'The sample size available for this study depended on a prospective cohort study, the RISE-UP study, which provided data on 450 patients (ref). For logistic regression analysis, at least 10 event per candidate predictor are needed, according to prediction modelling guidelines. We assumed that the mortality rate would be around 11%. Therefore, the inclusion of 450 patients provided more than sufficient observations for our primary objective'.

2) The authors have stated that the venous blood samples were spun down and saved for analysis several weeks or months later. Could the authors confirm if this is standard procedure and that this methodology does not affect the quality of the results? For example, was there a difference between the results measured using this procedure compared to the results obtained when the clinician treating the patient requested it in the clinical setting?

This procedure is according to the laboratory standards. A clinical chemist was involved in the study protocol and stability of the biomarkers was guaranteed within a predetermined time period. All biomarkers were analysed within this time period and hence, the quality of the results was not affected in a negative way.

3) In this study the authors have treated each biomarker independently and found that each had a 'sufficient' or 'good' AUC to predict mortality. Have the authors considered using the biomarkers as a panel and studied their collective predictive accuracy? If not, why not?

Thank you for this suggestion. We focused on the individual biomarkers in this study. However, we did study them as a panel using multivariable logistic regression analysis. We found that several biomarkers together had an AUC of 0.82 (95% CI: 0.76 – 0.87), which is not higher than the discriminatory ability of our recently developed RISE UP score (Zelis et. al. EJIM 2020). We agree that this information may be useful for the readers and therefore we added the following text to the manuscript:

The results of this subanalysis were added to the result section (p 11) as follows:

"Subanalysis of combining biomarkers

In order to assess the discriminatory value of multiple biomarkers we developed a model through backwards elimination in the multiple logistic regression analysis. PCT did not contribute significantly to the model (p-value 0.51) and was therefore removed (Table 4). This resulted in a model consisting

of lactate, hs-cTnT, NT-proBNP and D-dimer. The AUC for prediction of 30-day mortality of these four biomarkers combined was 0.82 (95% CI: 0.76 – 0.87)."

We also added the results to Table 4.

We added following to the discussion section (p 13):

"In the multivariable analysis, stepwise elimination resulted in a new model consisting of four biomarkers, lactate, hs-cTnT, NT-proBNP and D-dimer, which yielded an AUC of 0.82. This discriminatory ability was, however, not better than that of the recently developed RISE UP score (AUC 0.83) which consists of age, vital signs and four routine laboratory tests albumin, blood urea nitrogen, lactate dehydrogenase, and bilirubin [1]. The RISE UP score was developed in the same patient sample and has the advantage of using inexpensive variables, which are collected in routine ED care making the score feasible for use in older ED patients. In addition, we recently showed that adding these biomarkers to the RISE UP model only minimally improved the AUC of the model by 0.03 [2].The limited added discriminatory ability and the expected extra costs support our conclusion that routinely determined biomarkers are not beneficial for prediction of mortality in older ED patients."

..And...we added the following to the analysis section (p 8):

"As a subanalysis, we evaluated the discriminatory ability of a combination of biomarkers. For this purpose, we used logistic regression with backwards elimination using a p-value of 0.10 for removal and determined the discriminatory ability of this new model by calculating AUC with 95% CI."

4) If the collective panel of biomarkers is predictive of mortality, how does it compare with the new simplified model [Ref 17] they have created? Do any of the biomarkers either individually or collectively add value to their model in Ref 17?

This is a very important question. Please see our answer to question 3.

5) It is concerning that over 15% of patients were missing a lactate, when only 5-10% of other biomarkers were missing? Why is this?

Thie reason for this was that lactate was measured in arterial blood samples, while the other biomarkers were measured in venous samples. Since some patients refused arterial blood sampling, this resulted in more missing values of lactate.

6) The conclusion the authors have arrived at is correct based on the analysis they have carried out but I do feel the analysis is inadequate for reasons given in comments 3 & 4.

Thank you for this feedback. Please see our answer to question 3.

Minor concerns:
1) In the introduction section, use the term emergency department once before it is abbreviated to ED.

Indeed, we did not explain this term. We changed the introduction section (p 5) so this term is explained earlier.

2) In the methods section, the authors have stated that p-values ≤0.05 were considered significant. Is this a typographical error? It is usually p<0.05 that is considered significant.

Indeed, this is an typographical error. We changed it into: “p-values <0.05” (page 9, Analysis)

3) The flow chart for patient selection should be part of the main paper, not a supplementary file.

Thank you for this kind suggestion. We added the patient selection to the main paper (Figure 1, page 9, Results).

Reviewer: 2

Reviewer Name: Camille Schwab
Institution and Country:
INSERM, Institut Pierre Louis d’Epidémiologie et de Santé Publique, F75012, Paris, France ; Université Paris-Sud, Faculté de Pharmacie, Département de Pharmacie Clinique, F-92296, Châtenay-Malabry, France ; GHU APHP.Sorbonne Université, Hôpital Saint Antoine, Service Pharmacie, F75012, Paris, France
Competing interests: None declared

Please leave your comments for the authors below
Thank you for the opportunity to review this paper. I have some remarks and questions:

Thank you for your constructive feedback. We feel that the changes we made accordingly improved the manuscript substantially. Please find the answers to your remarks below.

In the introduction section:
Background
The authors wrote that the biomarkers are used to “estimate the severity of specific diseases”. But these biomarkers are first used to diagnose the diseases, and then the severity.

Thank you for this comment. We agree with you that the words used could have been more precise and therefore changed the sentence into: ‘Biomarkers such as lactate, high-sensitivity cardiac Troponin T (hs-cTnT), N-terminal pro-B-type natriuretic peptide (NT-proBNP), D-dimer and procalcitonin (PCT) are frequently used to **diagnose and** estimate the severity of specific diseases.’ (p 5, background)

Importance
Can the authors better explain why were they interested in biomarkers that are mostly present in acute diseases (heart failure, myocardial injury) and which have fast kinetics, for long term outcomes (30-day mortality or 30-day unplanned readmission)?
The scope of the RISE UP study was adverse outcomes within 30days after initial emergency department presentation. We consider these outcomes short-term outcomes. Because these 5 biomarkers are markers of underlying diseases with high mortality rates, we included them in our study. In addition, for diseases such as myocardial infarction, heart failure or sepsis, death does not always occur soon after the initial diagnosis. Furthermore, we chose this follow up period because we think 30-day mortality is important for clinical decision making in the ED and this time period is used in other ED risk stratification scores as well (e.g. CURB-65 score or abbMEDS [3 4]).

In the method section:

Study design, setting and selection of participants:
1) As this study is part of the RISE UP study, why did it take place only in the Zuyderland MD and not in all the centers of the RISE UP study? I think the authors should explain why.

Indeed, this was not explained in the main text. In the RISE UP study, additional blood samples were drawn in patients included in Zuyderland MC for feasibility reasons.

We changed the following sentence in the methods section, p 6:

"This part of the study took place in Zuyderland MC, a large teaching hospital in the south of the Netherlands, ...**because biomarkers were measured in patients included in this site only."**

In addition, we added a sentence to the methods section, p 6:

"The study protocol of this study was published online [5]."

2) Why did the patients were included only if they were treated by an internist or a gastroenterologist? I think it conducts to a classification bias as gastroenterologist's patients have very specific diseases.

In the Netherlands, internal medicine and gastroenterology are part of the same department and internal medicine and gastroenterology patients are treated by the same doctors in the ED. In addition, we think that both specialties treat the same groups of older patients with multimorbidity and complex acute medical problems (e.g. patients with pancreatitis, liver cirrhosis or cholangitis).

Measurements
3) Can the authors explain why did the samples were analyzed within 4 weeks or 3-4 months after collection?

This is a very good question. We aimed to blind the results of the biomarkers for the physicians in order to minimalize influence on usual medical care. In normal clinical practice, these biomarkers are not measured, unless on specific request of the physician. We therefore decided to analyze the biomarkers measured in venous blood samples in a later stage. Since lactate was immediately measured in an arterial blood gas for technical reasons, only the results of this analysis were available to the attending physician.

4) As the samples were analyzed 1 to 4 months after the ED visit, when did the data were collected from the EMR? Did the Charlson Index was calculated after the ED visit, and thus include the possible new comorbidity?

The data were collected at the moment of inclusion in the study. The charlson comorbidity index was calculated at the moment of presentation to the ED and did not include any new discovered comorbidity. Only the outcome measures were collected at a later stage.

We changed following sentence in the methods section:

"All were collected from electronical medical records. The following data were retrieved upon arrival at the ED visit: age, sex and living situation, data on comorbidity according to the Charlson Comorbidity

Index (CCI)[6] and triage category (using the Manchester Triage System (MTS)[7]." (page 6, measurements)

Outcome

5) Can the authors explain how they measured the outcome:

- How were the deaths verified?

We verified the deaths by checking the medical files, which are connected to the municipal administrational records, and, if data were inconclusive, by contacting the general practitioner (GP) for additional information.

We added following to the methods section:
"Data regarding the outcomes was collected by checking the electronical medical files, which are connected to the municipal administration and by contacting the general practitioner (GP) if necessary." (page 7, outcomes)

- Are the ICU/MCU admission had to be in the Zuyderland MD hospital, or in any another hospital, right after the ED visit or within 30 days after the ED visit?

ICU/MCU admission had to occur within 30 days after the ED visit. Zuyderland MC is a large teaching hospital, with a large ICU unit. It is highly unusual to admit patient to another hospital in the region than the one they initially visited. This is also described in detail in our study protocol, to which we refer in the methods.

(page 6, methods, design, see our response to your comment nr. 1 on the design).

- For the LOS, was it an admission in the Zuyderland MD hospital, or in any another hospital, right after the ED visit or within 30 days after the ED visit?

For the analysis of the length of hospital stay, we took the length of admission following the ED visit into account. Revisits were not included in this analysis.

We added following sentence to the methods section (section Outcomes, p7):

"Length of hospital stay was retrieved for patients who were admitted immediately following the ED visit".

- How was measured the loss of independent living?

Loss of independent living was defined as: discharge to a nursing home/hospice or with palliative care in previously community dwelling patients.

We added following sentence to the methods section (section Outcomes, p 7):

"Loss of independent living was defined as discharge to a nursing home/hospice or with palliative care in previously community dwelling patients."

- How was measured unplanned readmission? Was it unplanned readmission in Zuyderlann MD hospital, or in any another hospital?

Unplanned readmission was defined as a readmission to Zuyderland MC within the follow up period of 30 days after discharge from the ED or hospital. In the Netherlands, most patients are referred to the hospital by a general practitioner and return to this hospital when new medical problems arise or symptoms recur. We therefore included readmissions in Zuyderland MC only. Most patients were

followed up in the outpatients clinics and therefore an admission to another hospital would have been retrieved in their in the medical files. In addition, the GP was contacted when a patient did not return to the hospital afterwards.

Analysis

6) "Logistic regression analyses were performed on data after imputation of missing values". The authors have not conducted sensitivity analyses to make sure of the validity of the imputation.

Thank you for this comment. Although we did not state this in the manuscript file, we did perform sensitivity analyses to make sure that the choice of the method to handle missing data (in this case imputation versus complete case analysis) did not change the conclusion of our manuscript. For example, we checked the mean and median biomarker values before and after imputation, compared their AUCs and checked their predictive value in the logistic regression analysis.

We added following sentence in the analysis section: 'Sensitivity analyses were performed to assess the impact of missing data on our results by comparing the results after imputation to complete case analysis." (page 9, Analysis)

7) Contrary to what I read in the STROBE statements, I have not read in the manuscript how the study size was arrived at.

You are correct. We added a statement on the sample size calculation to the analysis part (page 8, Analysis). Please also see our answer to question 1 from reviewer 1.

8) "Linearity was visually checked", could the graphs be presented in the supplementary data?

We added the figures to the end of this rebuttal. Because logistic regression assumes linearity of independent variables and the log odds of experiencing the event of interest. We checked this before running the logistic regression analysis. In prediction modeling this makes sure the preditive ability of a predictor is optimal. In case a predictor is not linearly associated with the occurrence of an event, predictive performance will be lower. In such case, we would have considered categorizing a potential predictor. We don't think that the figures are useful to the readers but should you or the editor insist that they should be placed in the supplemental file, we are willing to do so.

In the results section:

Characteristics of study subjects

1) "During the follow up period", as I wrote earlier, can the authors explain in the method section how did they follow the patients?

Immediately after the ED visit, we collected the data of the ED visit. Thirty days after the ED visit and 30 days after hospital discharge, we checked all medial files which are connected to the municipal administrational records with respect to data of the outcomes. If data were missing, we contacted the general practitioner to obtain information regarding the outcomes.

We changed following sentence in the methods section (methods, measurements, p 7):
"All were collected from electronical medical records. The following data were retrieved upon arrival at the ED visit: age, sex and living situation, data on comorbidity according to the Charlson Comorbidity Index (CCI)[6] and triage category (using the Manchester Triage System (MTS)[7]."

We added following sentence to the methods section (outcomes section, p7):
"Data regarding the outcomes was collected by checking the electronical medical files, which are connected to the municipal administration and by contacting the general practitioner if necessary."

2) I think the chart inclusion is not complete. Can the authors add the outcomes in this flow chart?

Since the outcomes are part of the results we think that it is not correct to put these in the inclusion chart. If the reviewers or editor insist that we add this to the figure, we are absolutely willing to do so.

3) In the table 1, the authors have not mentioned the missing values. Can they add the percentage of missing values for each variable so that the reader can better understand results?

We added the percentage of missing values to table 1 and removed the column of totals.

4) "Non-survivors more frequently experienced the composite endpoint (72.5%)…": Can the authors had the N with the percentage, and had this interesting result in the table 1?

Thank you for this suggestion.

We added the results to table 1 and changed following sentence in the methods section (p 9):

"Non-survivors more frequently experienced the composite endpoint (n=37, 72.5%) compared to the survivors (n=164, 41.1%, p-value <0.001)."

(page 9, results, characteristics)

5) In the table 1, we can see that the survivors are very different from the non-survivors (age, community-dwelling, Charlson index), but the authors have not taken these differences into account for the logistic regression analyses, and presented non adjusted OR.

The scope of our study was to assess the predictive ability of the biomarkers for the outcomes. We did not intend to investigate the causal relationship between the biomarkers and the outcome. Therefore, no adjustments for variables that would be considered confounders in a causal analysis were performed. If we would have done so, the resulting predictive performance would not reflect the predictive performance of the biomarker, but the predictive performance of a model including the biomarker and a number of confounders. That would greatly overestimate the performance of a single biomarker for the prediction of events.

Main results

6) "The highest values of the biomarkers were more often….between the non-survivors and survivors": Have the authors statistically measured this result?

Yes, we did measure this since we compared the median values and IQRs of the biomarkers in non-survivors and survivors. We found higher median values (and IQRs) for the non-survivors and found these to be significantly higher in non-survivors as stated in Table 1.

We adjusted following sentence in the results section (p 10, biomarkers):

"The highest values of the biomarkers were more often present in non-survivors whereas the lowest values were more often present in survivors, but there was a large overlap between the non-survivors and survivors (Table 1, Figure 2)."

7) Have the authors evaluated the predictive validity of the biomarkers of each outcome of the composite outcome, individually?

We did not evaluate the outcomes separately because some outcomes occurred not often enough to make these separate analyses (for example, only a minority of patients was admitted to the ICU).

In the discussion section:
1) "We detected no evidence for selection bias but cannot exclude it either" How did the authors evaluated the absence of selection bias?

We compared the baseline characteristics of included patients of the RISE UP study with the first 200 non-included patients. We have presented the results of this comparison in two prior publications of the RISE UP study (Zelis et al., Plos One 2018[8] and Zelis et al., EJIM 2020[1]).

Below, please find the supplemental table of our publication in EJIM below:

Supplemental Table S2. *Baseline characteristics of the included and non-included patients*

Characteristic	Reference values	Included[a] (n = 603)	Non-included[a] (n = 200)
Median age (IQR), years		79 (73–85)	78 (72–85)
Male sex, n (%)		311 (51.6)	100 (50.0)
Living in nursing- or care home, n (%)		51 (8.5)	23 (11.7)
Manchester triage category, n (%)			
Orange		66 (11.1)[d]	38 (19.2)[d]
Yellow		334 (56.1)	108 (54.5)
Green		189 (31.8)	51 (25.8)
Abnormal vital signs, n (%)			
Tachycardia >90 bpm		211 (35.5)	68 (34.2)
MAP <70 mmHg		41 (7.0)[d]	24 (12.5)[d]
Hypo- or hyperthermia (<36 or >38°C)		154 (26.5)	58 (29.6)
Tachypnea >20/min		137 (22.7)	44 (22.0)
O2-saturation <95%		131 (22.4)	45 (23.2)
GCS <15		91 (15.1)	38 (19.0)
≥2 abnormal vital signs[b]		168 (28.7)	63 (32.1)
Laboratory results			

Mean sodium (SD) , mmol/L	135–145	137 (5.2)	138 (6.4)
Mean potassium (SD), mmol/L	3.5–4.5	4.2 (0.7)	4.3 (0.7)
Median BUN (IQR), mmol/L	3.5–7.5	8.0 (5.6–13.1)	9.0 (6.1–13.6)
Median creatinine (IQR), **µmol/L**	70–110	97 (75–137)	103 (82–145)
Mean calcium (SD), mmol/L[c]	2.20–2.60	2.37 (0.17)	2.40 (0.23)
Median bilirubin (IQR), **µmol/L**	<21	10.0 (6.0–15.0)	9.0 (5.0–15.0)
Median alkaline phosphatase (IQR), U/L	♀<98 ♂<115	87 (70–119)	83 (65–110)
Median GGT (IQR), U/L	♀<38 ♂<55	34 (20–71)	29 (21–66)
Median ALT (IQR), U/L	♀<34 ♂<45	21 (15–31)	22 (16–30)
Median AST (IQR), U/L	♀ <31 ♂<35	27 (20–43)	26 (20–43)
Median LDH (IQR), U/L	♀ <247 ♂<248	226 (186–295)[d]	216 (175–277)[d]
Mean albumin (SD), g/L	35–50	36.3 (5.7)	37.0 (5.0)
Median CRP (IQR), mg/L	<10	31 (7–88)	31 (4–80)
Treatment , n (%)			
Internal medicine		455 (75.5)	153 (76.5)
Admission		479 (79.4)	171 (85.5)
30-day mortality, n (%)		66 (10.9)	30 (15.0)

ALT, alanine aminotransferase; AST, aspartate aminotransferase; BUN, blood urea nitrogen; CRP, C-reactive protein; GCS, Glasgow Coma Scale; GGT, gamma-glutamyltransferase; IQR, interquartile range; LDH, lactate dehydrogenase; MAP, mean arterial pressure; NA, not applicable; SD, standard deviation.

To convert the laboratory tests values from SI-units to conventional units multiply by the following conversion factors: for sodium to mEq/L by 1.0, potassium to mEq/L by 1.0, BUN to mg/dL by 2.8, creatinine **to mg/dL by 0.0113, calcium to mg/dL by 4, bilirubin to mg/dL by 0.0585 and albumin to g/dL by 0.1.**

[a]In total, 603 patients were included in the derivation cohort during the prospective study period (included patients). The non-included patients represent a sample of 200 patients who were not included during the prospective study period for reasons other than refusal.

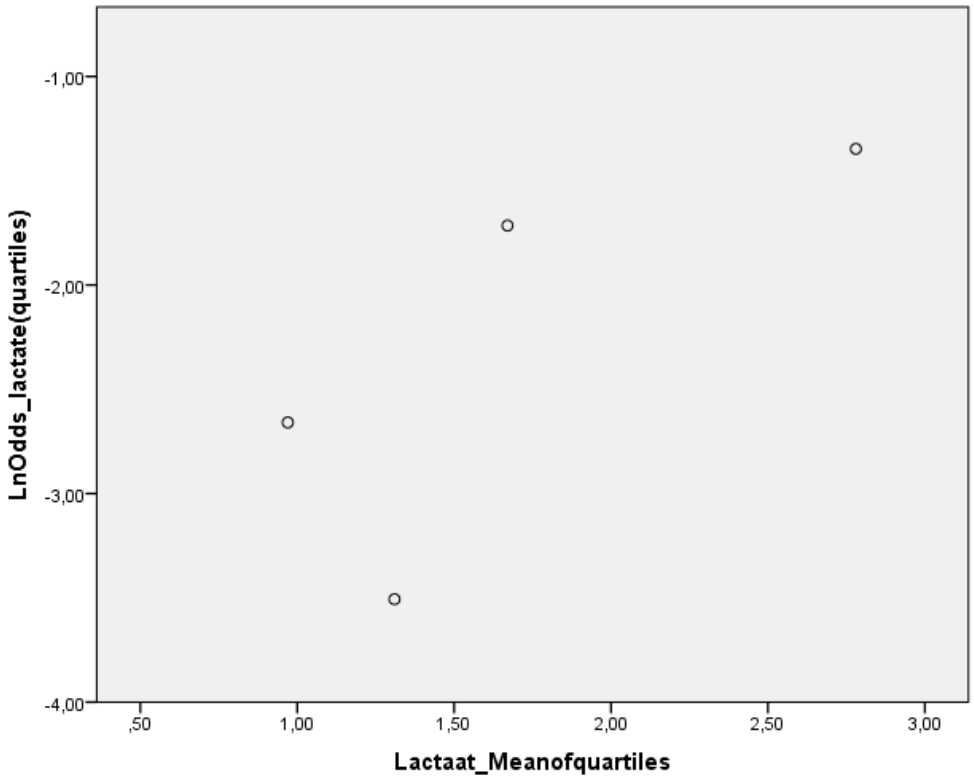
[b]≥2 of following abnormal vital signs: HR >90 bpm, MAP<70 mmHg, RR>20/min, O2-saturation<95% and GCS<15

[c]Calcium was corrected for albumin by the formula: *calcium + (0.02*(40-albumin))*.

[d]There were no significant differences in baseline characteristics between the two groups, except for triage category orange (P=0.003), MAP (P=0.02) and serum LDH (0.03).

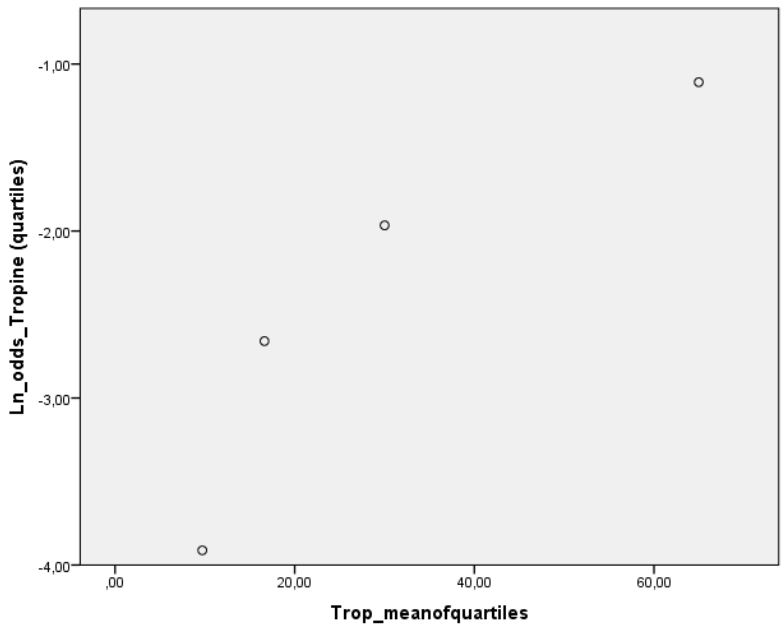
Additional Figures:

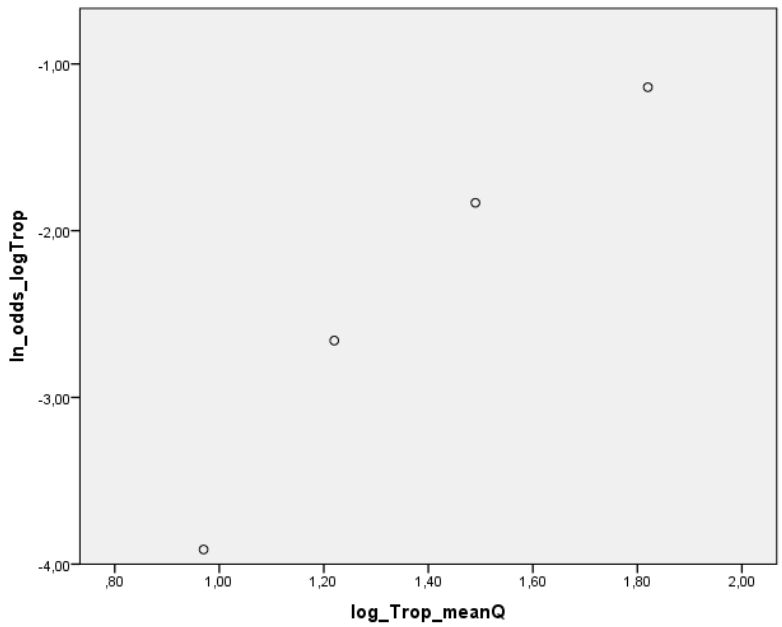
1) **Lactate**



Lactate: conclusion: not-linear nor logarithmic. Therefore, lactate was dichotomized.

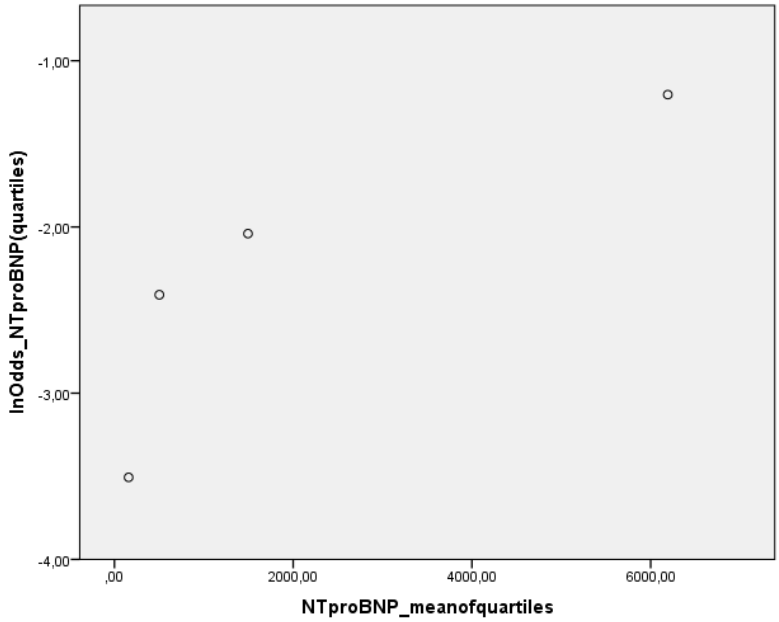
1) **Troponin**

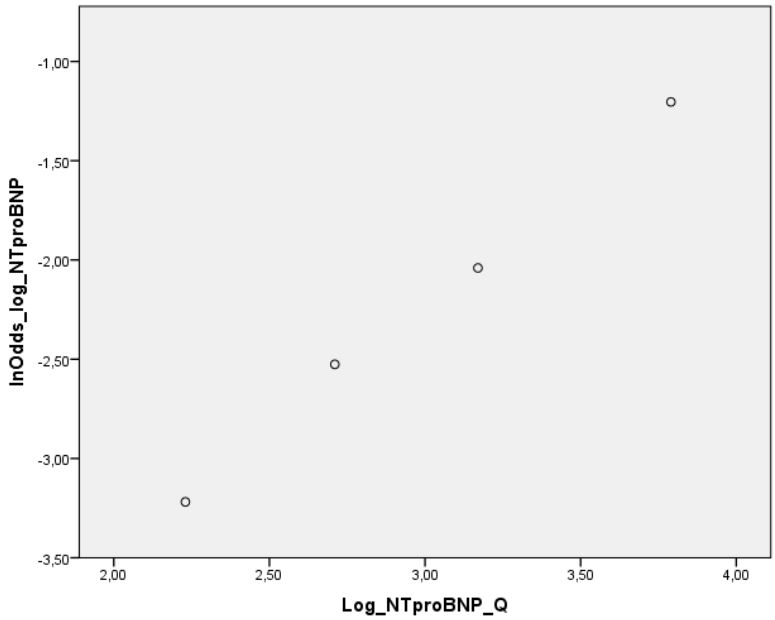




Troponin: conclusion not linear but logarithmic.

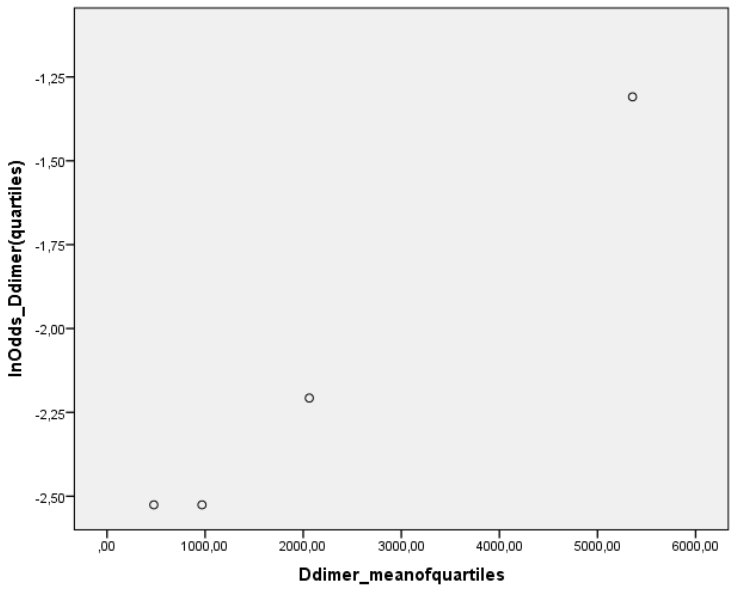
1) **NT-proBNP**





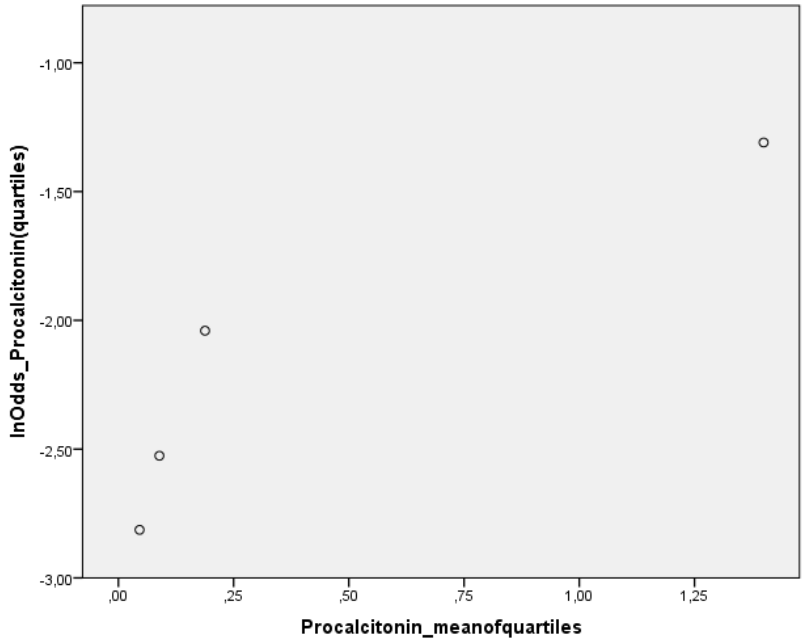
NT-proBNP: conclusion: not linear but logarithmic.

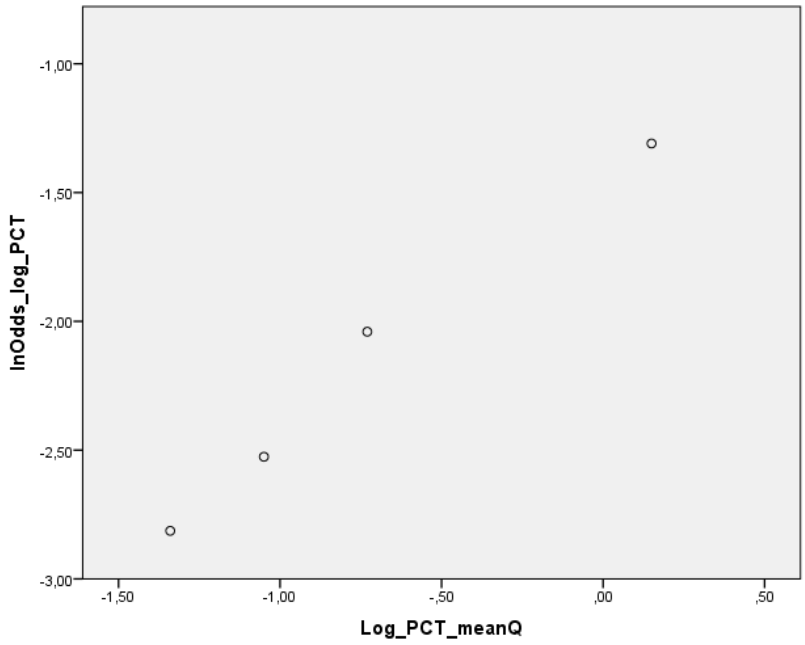
2) D-dimer



D-dimer: conclusion: not linear nor logarithmic. Therefore, d-dimer was dichotomized.

5) Procalcitonin





Procalcitonin: conclusion not linear but logarithmic.

References:

1. Zelis N, Buijs J, de Leeuw PW, et al. A new simplified model for predicting 30-day mortality in older medical emergency department patients: The rise up score. *European journal of internal medicine* 2020 doi: 10.1016/j.ejim.2020.02.021 [published Online First: 2020/03/03]
2. Zelis N, Buijs J, de Leeuw PW, et al. Do biomarkers add anything to clinical prediction of mortality in older medical emergency department patients? *European journal of internal medicine* 2020 doi: 10.1016/j.ejim.2020.07.016 [published Online First: 2020/07/29]
3. Vorwerk C, Loryman B, Coats TJ, et al. Prediction of mortality in adult emergency department patients with sepsis. *Emergency medicine journal : EMJ* 2009;26(4):254-8. doi: 10.1136/emj.2007.053298 [published Online First: 2009/03/25]

4. Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003;58(5):377-82. [published Online First: 2003/05/03]
5. Zelis N, Buijs J, de Leeuw PW, et al. Study protocol for a multicentre prospective cohort study to identify predictors of adverse outcome in older medical emergency department patients (the Risk Stratification in the Emergency Department in Acutely Ill Older Patients (RISE UP) study). *BMC geriatrics* 2019;19(1):65. doi: 10.1186/s12877-019-1078-2 [published Online First: 2019/03/06]
6. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40(5):373-83. [published Online First: 1987/01/01]
7. Mackway-Jones K. Emergency triage. London: BMJ Publishing Group 1997 1997.
8. Zelis N, Mauritz AN, Kuijpers LIJ, et al. Short-term mortality in older medical emergency patients can be predicted using clinical intuition: A prospective study. *PloS one* 2019;14(1):e0208741. doi: 10.1371/journal.pone.0208741 [published Online First: 2019/01/03]

VERSION 2 – REVIEW

REVIEWER	Narani Sivayoham St George's University Hospitals NHS FT Blackshaw Road London SW17 0QT United Kingdom
REVIEW RETURNED	25-Nov-2020

GENERAL COMMENTS	Thank you for answering my queries. Congratulations on your manuscript.

REVIEWER	Camille Schwab GHU AP-HP.Sorbonne Université, Hôpital Saint Antoine, Pharmacie Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F75012, Paris, France Université Paris-Saclay, Faculté de Pharmacie, Département de Pharmacie Clinique, F92296 Chatenay Malabry, France
REVIEW RETURNED	03-Dec-2020

GENERAL COMMENTS	The authors have responded to most of my previous remarks. While responses to all comments would have been appreciated, the authors have improved their article and made it acceptable for publication, in my opinion.