Biomarker validation in the emergency department. General criteria and clinical implications

Giuseppe Lippi,[1] Camilla Mattiuzzi,[2] Gianfranco Cervellin[3]

[1]Laboratory of Clinical Chemistry and Hematology, Pathology and Laboratory Medicine Department, Parma University Hospital; [2]Clinical Governance Service, Trento General Hospital; [3]Emergency Department, Parma University Hospital, Italy

Abstract

The routine use of biomarkers, which is also rapidly expanding in the emergency department, carries some potential drawbacks such as the risk of producing false positive results and also places a substantial economical burden on the healthcare system, especially when the use of laboratory resources is poorly discretionary or even inappropriate. The aim of this article is to provide an overview about some general criteria for biomarker validation in the emergency department, and discuss some relevant clinical implications. The leading aspects include analysis of data distribution and diagnostic performance, along with economical and organizational issues. We also brought a pragmatic example, comparing creatine kinase MB, a contemporary-sensitive troponin I and a high-sensitivity troponin I immunoassays for evaluation of patients with suspected acute myocardial infarction at emergency department admission.

Introduction

A biological marker, also conventionally known as *biomarker*, is typically defined as a characteristic that can be measured and used as an indicator of biology, of a pathological process, or even as a guide for targeting specific therapeutic interventions.[1,2] Several lines of evidence now attest that the routine use of biomarkers for screening, diagnosis, prognostication, therapeutic monitoring and follow-up of the vast majority of human disorders is virtually unalienable. The greater advantages of biomarker assessment entails the scarce invasivity wherein a collection of venous blood is only required for performing an impressive number of tests, the short turnaround time (most laboratory tests can be performed in less than 30 min), the possibility to process large volumes of analyses due to automation of testing, the relatively low cost as compared with other diagnostic investigations (*e.g.*, diagnostic imaging), and the objective interpretation of data since transversal (*i.e.*, against a specific reference range) or longitudinal (*i.e.*, against previous patient data) comparison of test results is an objective means for establishing whether a given parameter is diagnostic or not.

The routine use of biomarkers is rapidly expanding in all fields of science and medicine, including the emergency department (ED).[3-5] Incidentally, biomarker testing finds its natural application in this peculiar healthcare setting, because it provides a rapid means for assessing patients in a growingly overcrowded environment, the collection of blood specimens does not require specific skills, the performance and interpretation of test results does not conventionally require the intervention or assistance of other physicians as for radiological examinations.[6] In the ED, biomarkers can be used for achieving a final diagnosis of disease and, even most frequently, for ruling out a clinical suspicion and thereby allowing a safe discharge of patients.[7]

After that said, the increasingly use of biomarkers in the ED carries also some potential drawbacks, such as the risk of producing false positive results due to the statistics used for establishing reference ranges,[8] and also places a substantial economical burden on ED and laboratory, especially when the use of laboratory resources is poorly discretionary, or even inappropriate. Therefore, the aim of this article is to provide an overview about some general criteria for biomarker validation in the ED, and discuss some relevant clinical implications. For a pragmatic interpretation of several concepts, we will use the emblematic example of acute myocardial infarction (AMI) diagnostics at patient admission in the emergency room. This challenging paradigm is particularly suited for the topic of this article, since it represents the leading cause of ED admission, and the diagnostic work up is now largely dependent upon results of biomarker testing.[9,10] The data used in this article were obtained in a population of 98 consecutive patients (mean age 67 years, range: 35-84; 67 males and 31 females) admitted to the ED of the Academic Hospital of Parma for suspected AMI over three working days. Blood samples were collected at patient admission in primary blood tubes containing no additives (Becton Dickinson, Franklin Lakes, NJ, USA) and were rapidly transported to the central laboratory, where they were centrifuged at $1300 \times g$ for 10 min at room temperature. The serum was separated and stored in aliquots for further testing. For direct validation and comparison of diagnostic biomarkers, we decided to measure i) an *old* and virtually obsolete test such as creatine kinase isoenzyme MB (CK-MB; Beckman Coulter, Brea, CA, USA), which is characterized by a limit of detection (LOD) of 0.1 µg/L and a reference range comprised between 0.6 and 6.3 µg/L; ii) a contemporary-sensitive troponin I (TnI) test (AccuTnI; Beckman Coulter), which is characterized by a LoD of 10 ng/L and a 99th percentile of the upper reference limit (URL) of 56 ng/L; and iii) a high-sensitivity (HS) TnI test (HS-AccuTnI; Beckman Coulter), which represent the current gold standard for diagnosing myocardial injury according to most guidelines, and is characterized by a LoD of 2.5 ng/L and a 99th percentile of the URL of 32 ng/L.[11] All biomarkers were measured on the automated platform Access 2 (Beckman Coulter). A final diagnosis of AMI in our study population could be established in 11/98 patients (*i.e.*, 11%), according to the well established criteria of the third universal definition of myocardial infarction.[12]

The statistical analysis was performed with Analyse-it for Microsoft Excel (Analyse-it Software Ltd., Leeds, UK) and MedCalc Version 12.3.0 (MedCalc Software, Mariakerke, Belgium). All ED patients provided an informed consent for participating in this study, which was performed in agreement with the ethical standards established by the institution in which the experiments were performed and the Helsinki Declaration of 1975.

Opinion Report

Analysis of values distribution

The foremost step that must be undertaken before performing other types of statistical evaluation is to establish the pattern of data distribution (*i.e.*, Gaussian or not), which then

influences the selection of the most appropriate statistical approach. This can be accomplished with several tests, the most widely used are represented by Kolmogorov-Smirnov, D'Agostino-Pearson, Anderson-Darling and Shapiro-Wilk W tests. When the results of these tests (*i.e.*, the P value) are higher than 0.05, it can be assumed that data follow a normal (Gaussian) distribution, whereas in the presence of P values lower than 0.05 the hypothesis that the distribution of data is normal should be rejected.[13]

The analysis of our patient population, using the Kolmogorov-Smirnov test for normality, is shown in Figure 1, which clearly demonstrates that the values distribution of none of the three biomarkers follows a normal pattern, so that a non parametric approach must be used for their further analysis. This typically entails that patient values should be displayed in terms of median and percentile (or interquartile range; IQR) rather than in terms of mean and standard deviation, unless values are logarithmically transformed to achieve a normal distribution. Accordingly, the Wilcoxon-Mann-Whitney test (rather than the Student's *t* test, which is used for data normally distributed) is the preferred approach for distinguishing whether the results obtained in the populations with or without disease (*i.e.*, AMI) are significantly different. The results of biomarkers measurement in our study population are reported in Table 1, which shows that a significant difference exists between patients with or without a final diagnosis of AMI. This analysis is a necessary premise for establishing whether a biomarker may be clinically usable or not, because the lack of significant differences of values between patients with disease and those without would also imply a lack of real clinical usefulness. A graphical representation of data may also be useful under some circumstances, as shown in Figure 2, but this is not strictly necessary.

Receiver operating characteristic curve analysis

Once the analysis of raw data has been completed as described in the former paragraph, the second advisable step entails the evaluation of diagnostic performance of the test(s). The preferred means is indeed represented by the use of receiver operating characteristic (ROC) curve analysis. This approach allows to obtain a graphical representation of the overall diagnostic performance, which is typically quantified in terms of Area Under the Curve, AUC. The closest is the value of the AUC to 1.0, the better is the diagnostic performance. The output of the analysis is also supported by a degree of statistical significance (*i.e.*, P value) and a complete description of diagnostic sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), nega-

tive likelihood ratio (LR-) and positive likelihood ratio (LR+). The local calculation of these parameters, when feasible, is indeed preferable over the use of manufacturer's reference rages or predefined diagnostic thresholds established on the 99th percentile of the URL, since these limits have been calculated on populations that may largely differ from

those typically referred to the local ED.[14]

Indeed, ROC curves are very useful, but not completely pervasive in trials including a modest sample size. Nevertheless, this type of analysis is indeed the core for assessing the diagnostic performance of biomarkers and represents a suitable background for planning larger and more focused investigations, by
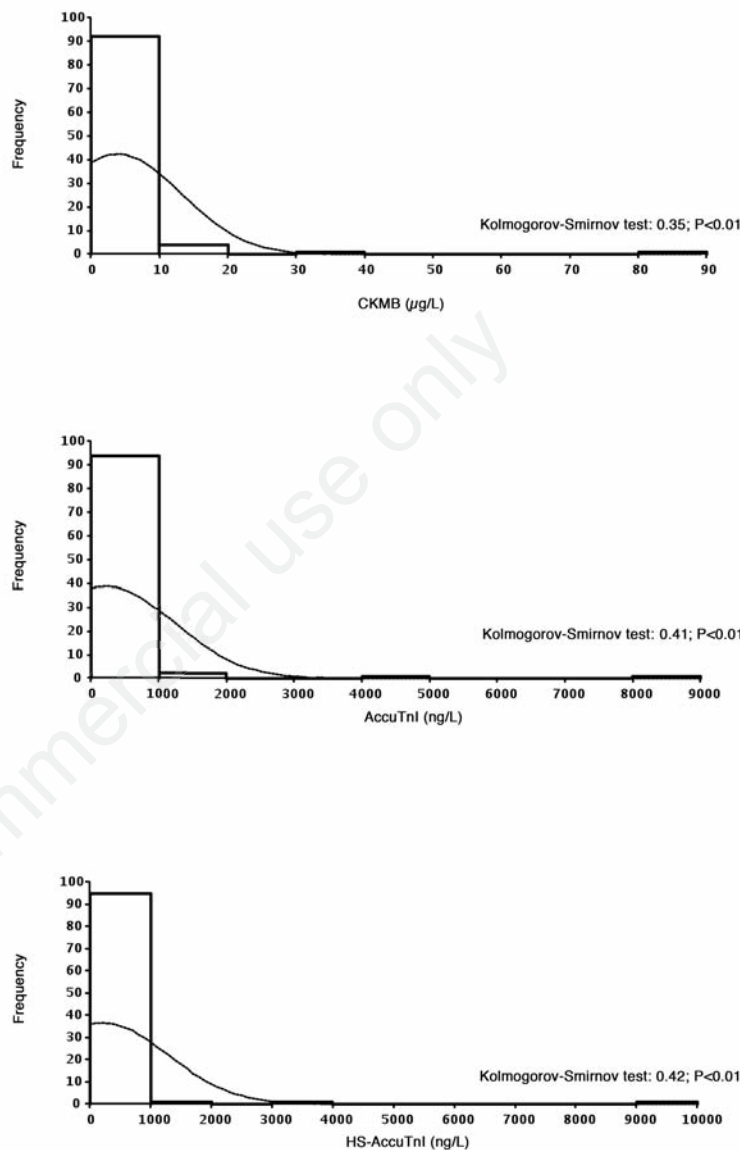

Figure 1. Value distribution of creatine kinase isoenzyme MB, a contemporary-sensitive Troponin I immunoassay and a high-sensitivity Troponin I immunoassay in a population of 98 consecutive patients admitted to the emergency department with suspected acute myocardial infarction. The assumption of normality is tested with Kolmogorov-Smirnov test.

Table 1. Results of biomarker measurement in a population of 98 consecutive patients admitted to the emergency department with suspected acute myocardial infarction.

	Non-AMI (n=87)	AMI (n=11)	P
CK-MB (µg/L)	1.9 (1.2-2.8)	7.5 (5.5-16.7)	<0.001
AccuTnI (ng/L)	3 (0-20)	900 (35-1373)	<0.001
HS-AccuTnI (ng/L)	7 (4-23)	621 (41-1237)	<0.001

AMI, acute myocardial infarction; CK-MB, creatine kinase isoenzyme MB; AccuTnI, contemporary-sensitive Troponin I immunoassay; HS-AccuTnI, high-sensitivity Troponin I immunoassay. Results are shown as median and interquartile range and the significance of differences is assessed with Wilcoxon-Mann-Whitney test.

determining decision thresholds and preliminary clinical usefulness.[15] According to the diagnostic performance, ROC curves thus allow to identify the *optimal* cut-off, that is the biomarker value associated with the highest result of the sum of sensibility and specificity. However, this value does not necessary represent the *best* threshold according to the intended clinical use of the test. The assessment of D-dimer for ruling out venous thromboembolism (VTE) is a typical example. According to its validated use, D-dimer testing typically follows the evaluation of pre-test probability of VTE, and its value influences the decision to performed additional imaging investigations (*e.g.*, leg ultrasonography or lung computed tomography).[16] As such, the best cut-off in this setting is not the compromise between sensitivity and specificity, but rather the D-dimer value characterized by the highest sensitivity (advisably 1.0), since this would allow to avoid additional (invasive) testing and discharge patients. Similar considerations can be drown for the use of proteins S100B for the diagnosis of brain injury in patients with mild head trauma, wherein a non diagnostic value of this biomarker should permit to prevent unnecessary head CT.[17]

The comparison of the ROC curves of the three biomarkers in our study population is shown in Figure 3. The AUC of the different biomarkers was 0.90 (95% CI, 0.77 to 1.00; P<0.001) for CK-MB, 0.92 (95% CI, 0.85 to 0.99; P<0.001) for AccuTnI and 0.92 (95% CI, 0.84 to 0.99) for HS-AccuTnI. Interestingly, although the AUCs of both TnI immunoassays were better than that of CK-MB, no significant differences were found (*i.e.*, CK-MB *vs* AccuTnI, P=0.68; CK-MB *vs* HS-AccuTnI, P=0.69; AccuTnI *vs* HS-AccuTnI, P=0.77). The optimal thresholds, which incidentally corresponds to the best cut-offs for both TnI immunoassays (*i.e.*, those associated with 1.00 sensitivity for rule out of AMI upon ED admission), were 4.6 µg/L for CK-MB, 17 ng/L for AccuTnI, and 14 ng/L for HS-AccuTnI, respectively (Table 2). Although the diagnostic performance of CK-MB was relatively poor as compared with those of both TnI immunoassays (*i.e.*, the specificity was higher but the sensitivity was indeed unacceptable), it is noteworthy that, in analogy with recently published data,[18,19] the diagnostic performance of the contemporary-sensitive (*i.e.*, AccuTnI) and HS (*i.e.*, HS-AccuTnI) tests was nearly identical, with a negligible better specificity and PPV of the former method. An additionally useful test is then represented by the calculation of the diagnostic odds ratio (DOR), which is synthetically defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease [*i.e.*, (true positive/false negative)/(false positive/ true negative)]. Also in this circumstance the DORs
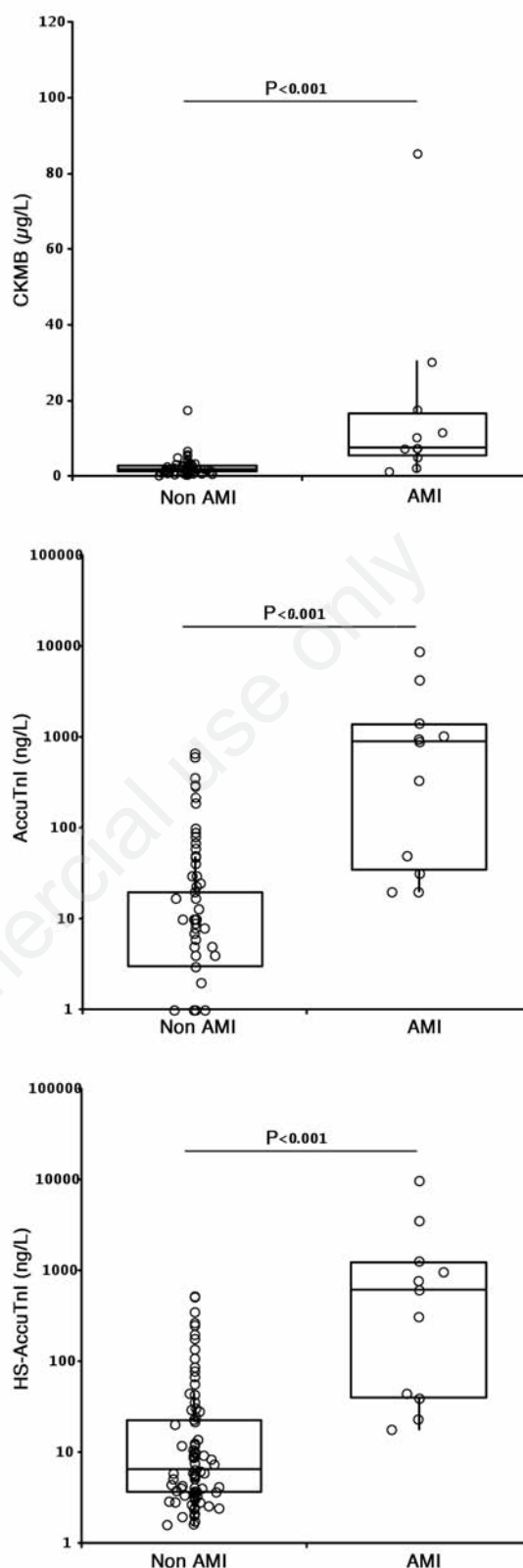

Figure 2. Value distribution of creatine kinase isoenzyme MB, a contemporary-sensitive Troponin I immunoassay and a high-sensitivity Troponin I immunoassay in a population of 98 consecutive patients admitted to the emergency department with suspected acute myocardial infarction. Results are shown as median and interquartile range and the significance of differences is assessed with Wilcoxon-Mann-Whitney test.

of both TnI immunoassays calculated in our study population outperform that of CK-MB, with the DOR of the contemporary assay being slightly higher than that of the HS due to a lower number of false positive cases (Table 2).

Organizational and cost-effective-ness analysis

A final important evaluation, especially in a world of limited resources, is the organizational and economical impact that the assessment of the various biomarker poses on the healthcare system and the potential savings due to early or more accurate diagnosis.[20,21]

As regards practical considerations, the turnaround time is the leading aspect to be considered. According to recent guidelines, which entails serial sampling of cardiac biomarkers at 0, 3 and 6 h,[22] it is absolutely necessary to use analytical techniques that can produce results in less than 60 min, which should also be located at a convenient distance from the ED to avoid delays due to sample transportation.[23] When both requisites cannot be fulfilled, introduction of point of care testing in the ED may be a viable option.[24]

As regards the economical burden, there is no simple means for calculating reliable figures, but one suitable approach can be suggested. The number needed to test (NNT) can be synthetically described as the number of patients that should be tested to identify one additional adverse outcome (*e.g.*, AMI). Although this calculation is more typically used for defining the number needed to screen (NNS) within policies of risk reduction (*e.g.*, prostate specific antigen screening for preventing mortality from prostate cancer), it can also be reliably applied for gathering information about the organizational and clinical burdens that one given test poses on healthcare resources.[25] It is rather obvious that the higher is the value of the NNT, the larger is the cost for diagnosing a given disorder. According to our data, the NTT is expectedly lower for both TnI immunoassays (*i.e.*, 9) as compared with CK-MB (*i.e.*, 11). By translation of this concept into economical terms (*i.e.*, the cost of a single determination is €0.92 for CK-MB and €1.12 for AccuTnI in our institution), the final cost of routinely using CK-MB or AccuTnI would result to be nearly identical (*i.e.*, €10.12 versus €10.08) (Table 3). Nevertheless, the cost of CK-MB will be then inevitably inflated by the larger expenditure that emerges from the considerable number of missed diagnoses upon patient admission (2/11; *i.e.*, 18% versus 0/11 with both AccuTnI tests). This cost, which cannot be reliably estimated, is mostly attributable to delayed therapeutic intervention, worse outcomes, longer stay and greater consumption of ED resources.
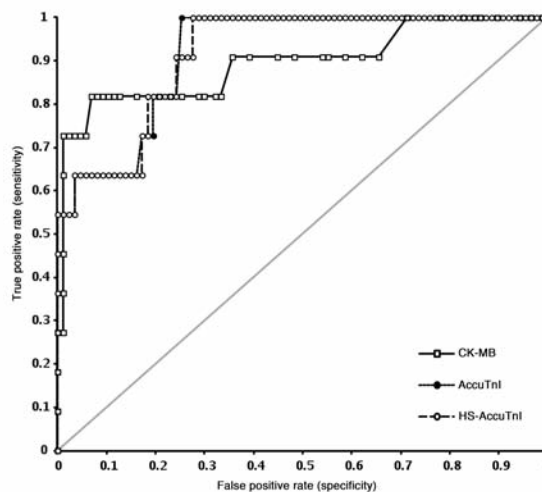

Figure 3. Receiver operating characteristics curve of creatine kinase isoenzyme MB, a contemporary-sensitive Troponin I immunoassay and a high-sensitivity Troponin I immunoassay for diagnosing acute myocardial infarction in a population of 98 consecutive patients admitted to the emergency department.

Table 2. Diagnostic performance of creatine kinase isoenzyme MB, a contemporary-sensitive Troponin I immunoassay and a high-sensitivity Troponin I immunoassay for diagnosing acute myocardial infarction in a population of 98 consecutive patients admitted to the emergency department.

	CK-MB	AccuTnI	HS-AccuTnI
AUC	0.90*	0.92°	0.92#
Cut-off	4.6 µg/L	17 ng/L	14 ng/L
Sensitivity	0.82	1.00	1.00
Specificity	0.93	0.75	0.72
NPV	0.98	1.00	1.00
PPV	0.60	0.33	0.31
LR-	0.20	0.00	0.00
LR+	11.86	3.95	3.63
DOR	51§	60^	56$

CK-MB, creatine kinase isoenzyme MB; AccuTnI, contemporary-sensitive Troponin I immunoassay; HS-AccuTnI, high-sensitivity Troponin I immunoassay; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; LR-, negative likelihood ratio; LR+, positive likelihood ratio; DOR, diagnostic odds ratio. *95% CI, 0.77 to 1.00; P<0.001; °95% CI, 0.85 to 0.99; P<0.001; #95% CI, 0.84 to 0.99; P<0.001; §95% CI, 9 to 286; P<0.01; ^95% CI, 3 to 1050; P<0.01; $95% CI, 3 to 993; P<0.01.

Table 3. Economical analysis of assessing creatine kinase isoenzyme MB, a contemporary-sensitive Troponin I immunoassay and a high-sensitivity Troponin I immunoassay for diagnosing acute myocardial infarction in a population of 98 consecutive patients admitted to the emergency department.

	CK-MB	AccuTnI	HS-AccuTnI
Number needed to test	11	9	9
Cost per test (€)	0.92	1.12	-
Total cost (€)	10.12	10.08	-

CK-MB, creatine kinase isoenzyme MB; AccuTnI, contemporary-sensitive Troponin I immunoassay; HS-AccuTnI, high-sensitivity Troponin I immunoassay.

Conclusions

Biomarkers validation is a challenging but necessary enterprise in almost every area of medical sciences, thus including the ED. The indiscriminate introduction of biomarkers in the ED, along with their inappropriate request, carries a high risk of consuming valuable human or economical resources and jeopardizing patient safety. In this article we have discussed some relevant issues that should guide the decision as to whether the introduction (or replacement) of a given biomarker is clinically justified and economically acceptable. According to data obtained using three different biomarkers (*i.e.*, *old*, *current* and *innovative*) for diagnosing AMI in the ED, we could confirm that the use of CK-MB appears now largely unjustified. It is noteworthy, however, that we could not find a real clinical improvement using a HS-TnI immunoassay as compared with the previous contemporary-sensitive method, at least upon patient admission to the ED. Although recent data attests that the former test would outstrip the previous techniques during serial testing according to its improved analytical sensitivity, this paradigm has recently been challenged by additional data showing that the diagnostic accuracy of some contemporary sensitive and HS immunoassays may be virtually identical using the recommended 2-3 h sampling protocol.[18,26] Despite the fact that we could not provide definitive economical data about the comparison of AccuTnI versus HS-AccuTnI because the latter test has not become commercially available so far, it is plausible to assume that the expenditure per test of the HS method would be not less than 20% higher (also also confirmed by the manufacturer). As such, considering that the clinical performance are virtually identical, it should not be ignored that the replacement of a well-suited contemporary-sensitive immunoassay with a novel HS method would be associated with a larger expenditure for the healthcare system. A final mention deserves the interface between the emergency physician, the laboratory professional and the medical direction. These three parties, which are actively involved in the project of introducing a novel biomarker, should actively cooperate for defining clinical paths, identifying the appropriate settings of implementation, as well as for monitoring organizational, clinical and economical outcomes. Active and forthright collaboration is the key to foster successful relationships and improve efficacy and efficiency in the ED.[27]

References

1. Lippi G, Plebani M. Biomarker research and leading causes of death worldwide: a rather feeble relationship. Clin Chem Lab Med 2013:51;1691-3.
2. Braunwald E. Biomarkers in heart failure. N Engl J Med. 2008;358:2148-59.
3. Di Somma S, Magrini L, Travaglino F, et al. Opinion paper on innovative approach of biomarkers for infectious diseases and sepsis management in the emergency department. Clin Chem Lab Med 2013;51: 1167-75.
4. Lippi G, Plebani M, Di Somma S, et al. Considerations for early acute myocardial infarction rule-out for emergency department chest pain patients: the case of copeptin. Clin Chem Lab Med 2012;50:243-53.
5. Lippi G, Valentino M, Cervellin G. Laboratory diagnosis of acute pancreatitis: in search of the Holy Grail. Crit Rev Cl Lab Sci 2012;49:18-31.
6. Cavazza M. Biomarkers in emergency medicine: great opportunities or expensive puzzles? Emerg Care J 2012;8:3-4.
7. Lippi G, Guidi GC. The power of negative thinking. Am J Emerg Med 2008;26:373-4.
8. Lippi G, Plebani M. False myths and legends in laboratory diagnostics. Clin Chem Lab Med 2013;51:2087-97.
9. Lippi G, Franchini M, Cervellin G. Diagnosis and management of ischemic heart disease. Semin Thromb Hemost 2013;39:202-13.
10. Lippi G, Cavazza M, Peracino A, et al. Ischemic heart disease in the emergency room: state of the art, innovation and research. Emerg Care J 2013;9:e7.
11. Apple FS, Ler R, Murakami MM. Determination of 19 cardiac troponin I and T assay 99th percentile values from a common presumably healthy population. Clin Chem 2012;58:1574-81.
12. Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction. Circulation 2012;126:2020-35.
13. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. Ann Intern Med 1981;94:555-600.
14. Lippi G, Margapoti R, Aloe R, Cervellin G. Highly-sensitive troponin I in patients admitted to the emergency room with acute infections. Eur J Intern Med 2013; 24:e57-8.
15. Kampfrath T, Levinson SS. Brief critical review: statistical assessment of biomarker performance. Clin Chim Acta 2013;419: 102-7.
16. Lippi G, Franchini M, Targher G, Favaloro EJ. Help me, Doctor! My D-dimer is raised. Ann Med 2008;40:594-605.
17. Cervellin G, Benatti M, Carbucicchio A, et al. Serum levels of protein S100B predict intracranial lesions in mild head injury. Clin Biochem 2012;45:408-11.
18. Lippi G, Cervellin G. Do we really need high-sensitivity troponin immunoassays in the emergency department? Maybe not. Clin Chem Lab Med 2013;52:205-11.
19. Lippi G, Cervellin G. Highly-sensitive immunoassays in the emergency department: counterpoint. Emerg Care J 2013;9: e16.
20. Lippi G, Cervellin G. Letter to the Editor: choosing troponin immunoassays in a world of limited resources. J Am Coll Cardiol 2013;62:647-8.
21. Iannone P. The need of a health technology assessment perspective in emergency medicine. Emerg Care J 2013;9:e8.
22. Casagranda I, Cavazza M, Clerico A, et al. Proposal for the use in emergency departments of cardiac troponins measured with the latest generation methods in patients with suspected acute coronary syndrome without persistent ST-segment elevation. Clin Chem Lab Med 2013;51:1727-37.
23. Lippi G, Simundic AM, Plebani M. Phlebotomy, stat testing and laboratory organization: an intriguing relationship. Clin Chem Lab Med 2012;50:2065-8.
24. Lippi G, Mattiuzzi C, Cervellin G. Point of care troponin testing: rules and regulations. J Electrocardiol 2013;46:727-8.
25. Rembold CM. Number needed to screen: development of a statistic for disease screening. Brit Med J 1998;317:307-12.
26. Lippi G, Cervellin G. Challenges of serial troponin testing: an unfinished symphony. Int J Cardiol 2013;168:4397.
27. Cavazza M. ECJ: a new journal for new challenges in emergency medicine. Emerg Care J 2013;9:e1.