

Biomarqueurs en pratique quotidienne : utiles et futiles

Bruno Riou

Service d'Accueil des Urgences, CHU Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, Sorbonne Universités, Université Pierre et Marie Curie-Paris 6, UMRS INSERM 1166, Paris, France

Correspondance: Pr. Bruno Riou, Service d'Accueil des Urgences, CHU Pitié-Salpêtrière, 47-83 Boulevard de l'Hôpital, 75651 Paris Cedex 13, France. Email : bruno.riou@psl.aphp.fr

Points essentiels

- Le biomarqueur est utilisé pour prédire un diagnostic, évaluer une gravité, ou prédire une réponse (efficacité, toxicité, pharmacocinétique) thérapeutique (théranostique) ou enfin indiquer un mécanisme physiopathologique sous-jacent...
- Les recommandations STARD (Standards for Reporting of Diagnostic Accuracy) ont été élaborées pour améliorer la qualité des publications dans le domaine des études diagnostiques.
- La mesure du biomarqueur est associée à une incertitude de mesure analytique. Aussi la précision de cette mesure doit être donnée (reproductibilité) ainsi que la limite de détection.
- La performance diagnostique d'un biomarqueur peut varier en fonction de la population étudiée et de ces caractéristiques démographiques (age, sexe) ou pathologiques (sévérité de la maladie, formes cliniques, sélection)

- les principes fondamentaux de la démarche scientifique expérimentale doivent être suivis : établir une hypothèse quantifiable testable (« réfutable » au sens de Karl Popper) et la tester.
- Toute étude diagnostique devrait comporter un calcul a priori du nombre de patients à inclure, calcul qui découle très logiquement de l'hypothèse que l'on souhaite tester.
- Lorsqu'un biomarqueur fait l'objet d'une étude, une question importante se pose à l'investigateur : quelle est la validité de l'étude.
- Etudes interventionnelles C'est un sujet rarement abordé dans la littérature et en pratique quotidienne, considérant de manière erronée que la prescription d'un biomarqueur ne peut avoir que des effets bénéfiques.
- Il y a donc un besoin urgent d'améliorer nos méthodes d'analyse des performances diagnostiques des biomarqueurs, particulièrement en ce qui concerne l'utilisation des courbes ROC, le choix des seuils, la définition de zone grise, le calcul du nombre de patients à inclure, et les techniques de validation interne. Les études sur les biomarqueurs peuvent présenter des limitations méthodologiques très importantes : les investigateurs doivent les prendre en compte dans la construction de leurs études, les rédacteurs en chefs et les relecteurs des journaux quand ils les analysent et les acceptent pour publication, et les lecteurs quand ils en prennent connaissance.

1. Introduction

Le terme de biomarqueur est utilisé pour n'importe quelle analyse biologique, incluant les analyses génomiques ou protéomiques, susceptibles de prédire un diagnostic, d'évaluer une gravité ou de prédire une réponse (efficacité, toxicité, pharmacocinétique) thérapeutique (théranostique) ou enfin indiquer un mécanisme physiopathologique sous-jacent [1]. De nouveaux biomarqueurs sont actuellement étudiés par les industriels de la biotechnologie et nous assistons à une révolution des biomarqueurs comme nous avons vécu une révolution de l'imagerie médicale [2]. Bien que des progrès importants aient été accomplis dans la façon dont les études diagnostiques doivent être rapportées [3], un long chemin reste à faire comme le montre la qualité méthodologique souvent médiocre des études diagnostiques publiées, y compris dans les revues les plus prestigieuses [4].

Le but de cet article est de décrire les principales recommandations pour l'établissement d'un protocole de recherche et la rédaction de la publication lors de l'étude d'un biomarqueur, permettant ainsi au lecteur d'articles scientifiques d'identifier les critères de qualité des publications dans ce domaine. Le lecteur trouvera des informations complémentaires sur les méthodes statistiques dans une revue générale récente [5]. Ce texte constitue une actualisation d'une publication récente [6].

2. A quoi sert un biomarqueur ?

Un biomarqueur peut jouer plusieurs rôles (Tableau I). Un biomarqueur peut être un marqueur diagnostic ou de sévérité d'une maladie. La troponine Ic est ainsi un marqueur sensible et spécifique d'infarctus du myocarde mais seulement un marqueur de sévérité (et non de diagnostic) de l'embolie pulmonaire alors que la procalcitonine est à la fois un marqueur diagnostique et de la sévérité de l'infection. Les biomarqueurs sont également utilisés pour la stratification, comme le lactate dans le sepsis ou la traumatologie. Toutefois, les objectifs diffèrent. Dans le diagnostic, l'évènement (maladie ou non) existe déjà, alors que dans le pronostic il s'agit d'un évènement à venir et il faut tenir compte de l'incertitude liée à sa réalisation ou non.

Plusieurs étapes d'importance croissante président à la démonstration de l'intérêt clinique d'un biomarqueur:

- démontrer que le biomarqueur est significativement modifié chez les malades par rapport aux témoins non malades;
- évaluer les propriétés diagnostiques du biomarqueur en les comparant à celles de la méthode de référence;

- démontrer que les propriétés diagnostiques du biomarqueur augmentent la capacité du médecin à prendre une bonne décision; ceci peut être parfois difficile car le moment du diagnostic peut être difficile à identifier ;
- évaluer l'utilité clinique d'un biomarqueur, à distinguer de l'évaluation de ses performances diagnostiques [7]. L'utilité peut comprendre des caractéristiques intrinsèques du test telles que le coût, son caractère invasif, les difficultés techniques, la rapidité d'obtention du résultat, et le contexte clinique (prévalence et pronostic de la maladie, coût et conséquences des traitements) ;
- démontrer que le biomarqueur modifie le pronostic dans le cadre d'études interventionnelles. Plusieurs études ont ainsi démontré que la procalcitonine permet de réduire la prescription d'antibiotiques chez les patients suspects d'une infection pulmonaire [8]. Toutefois, les études interventionnelles manquent pour la plupart des biomarqueurs [9].

Pour toutes ces étapes, il est important de comprendre les mécanismes physiopathologiques impliqués dans la synthèse et la libération du biomarqueur, sa cinétique et ses effets physiologiques propres. Le BNP est ainsi sécrété principalement par les myocytes du ventricule gauche en réponse à une augmentation de la tension pariétale, une expansion volumique, et une surcharge ventriculaire. Ses effets physiologiques comprennent une vasodilatation systémique et pulmonaire, un effet diurétique et natriurétique, une inhibition du système rénine-angiotensine-aldostérone et de l'endothéline. En revanche, les effets physiopathologiques de la procalcitonine restent mal connus.

Un biomarqueur peut également servir à guider d'autres types de décisions cliniques, particulièrement la prescription médicamenteuse. Ce domaine est particulièrement développé en oncologie où des biomarqueurs sont utilisés pour prédire l'efficacité ou la toxicité des chimiothérapies [1]. La procalcitonine a été utilisée pour guider l'arrêt de l'antibiothérapie en réanimation [10] et la détermination du phénotype de l'activation métabolique du clopidogrel pourrait améliorer le pronostic des patients ayant eu un infarctus du myocarde. Un biomarqueur pourrait également être utilisé comme critère de jugement de substitution [1,11] dans les essais cliniques.

3. Les recommandations STARD

Les recommandations STARD (Standards for Reporting of Diagnostic Accuracy) ont été élaborées pour améliorer la qualité des publications dans le domaine des études diagnostiques [3]. Un rapport scientifique complet et précis permet au lecteur de détecter des biais potentiels et de juger de l'applicabilité clinique et de la généralisation possible des résultats. Ces recommandations font suite aux recommandations CONSORT pour les essais cliniques

randomisés [12] et tentent de cerner les éléments qui peuvent limiter la validité interne ou externe des résultats obtenus (construction de l'étude, sélection des patients, méthode de référence, analyse des données). De nombreux journaux, parmi les plus prestigieux, demandent maintenant de se conformer à ces recommandations dans leurs instructions aux auteurs, et proposent même parfois de remplir une checkliste permettant aux auteurs de vérifier qu'ils ont effectivement satisfait aux critères demandés. Un audit des études diagnostiques publiées dans les meilleurs journaux médicaux de 1978 à 1993 a permis de constater que la méthodologie utilisée était souvent médiocre et que des éléments clés n'étaient le plus souvent pas décrits [13]. Des constatations similaires ont été faites dans les journaux spécialisés [14].

Les recommandations STARD (Table 2) proposent une liste de 25 items pour vérifier que l'information essentielle est bien décrite. L'utilisation d'un diagramme de sélection de la population étudiée (« flow chart ») est recommandée comme pour les essais cliniques, ainsi que l'utilisation généralisée des intervalles de confiance. Bien qu'il s'agisse d'une étape essentielle pour l'amélioration de la recherche dans les études diagnostique, il convient de souligner que les recommandations STARD abordent relativement peu la méthodologie statistique qu'il convient d'utiliser. Par ailleurs ces recommandations ne soulignent pas la nécessité d'élaborer les hypothèses de travail, d'écrire le protocole détaillé, y compris le plan d'analyse statistique et de le publier comme on le fait pour un essai clinique. Dans ce cadre l'identification de l'hypothèse testée et du critère de jugement principal est essentielle.

4. Les pièges de l'évaluation d'un biomarqueur

La mesure du biomarqueur est associée à une incertitude de mesure analytique. Aussi la précision de cette mesure doit être donnée (reproductibilité) ainsi que la limite de détection. La mesure d'un biomarqueur doit être sensible (détecter des concentrations faibles) et spécifique (sans interférence avec d'autres molécules surtout les métabolites). Toutes ces caractéristiques analytiques sont importantes et doivent être précisées car toutes les méthodes et toutes les trousse de dosage d'un même biomarqueur (par exemple la troponine) ne sont pas équivalentes entre elles. En particulier, les méthodes de biologie délocalisée ont souvent des caractéristiques analytiques moins bonnes. Les caractéristiques analytiques doivent être différenciées des caractéristiques diagnostiques [15]. Les termes « limite de détection » et « concentration minimale détectée » sont des synonymes pour exprimer la sensibilité analytique. La PCR (Polymerase Chain Reaction) est considérée comme une technique particulièrement sensible puisque capable de détecter quelques copies de gènes ou fragment de gènes. Malgré cette sensibilité analytique exceptionnelle, sa sensibilité diagnostique peut

être prise en défaut si le gène est absent du matériel examiné : c'est le cas d'un patient ayant une endocardite bactérienne mais dont l'hémoculture prélevée ne contient aucune bactérie. A l'inverse, la PCR peut être prise en défaut sur le plan diagnostique facilement par une contamination infinitésimale.

La performance diagnostique d'un biomarqueur peut varier en fonction de la population étudiée et de ces caractéristiques démographiques (age, sexe) ou pathologiques (sévérité de la maladie, formes cliniques, sélection) [16]. Un biomarqueur peut être intéressant dans une population donnée et non sur un sous groupe particulier. La procalcitonine est considérée comme un médiocre biomarqueur d'infection en cas de pyélonéphrite ou d'abcès intra-abdominal et n'a guère d'utilité dans une population exposée à un coup de chaleur. Les valeurs seuils de la troponine Ic pour diagnostiquer un infarctus myocardique postopératoire sont radicalement différentes en chirurgie cardiaque parce que celle-ci s'accompagne d'un relargage de troponine (traumatisme chirurgical, effet de la circulation extra-corporelle). C'est la raison pour laquelle il est important que les caractéristiques de la population étudiée soient bien précisées et expliquées aux lecteurs (Table 2).

Le problème des populations peut être élargi au problème plus général des influences extérieures (covariables). C'est le cas lorsque des facteurs sont susceptibles d'influer la qualité de la mesure (appareil, effet centre) ou la valeur du biomarqueur (cinétique). L'ajustement sur certaines covariables peut ainsi constituer un élément important de l'évaluation d'un biomarqueur [17].

L'étendue des valeurs de sensibilité et de spécificité rapportée pour un biomarqueur donné dans la littérature est souvent très large. Cette variabilité est sous-estimée dans la plupart des méta-analyses conduites [18]. En dehors du fait que les seuils utilisés dans les différentes études sont très variables, la raison la plus importante est que de nombreux biais existent [19] :

- la principale difficulté est la méthode de référence utilisée. Dans la plupart des études, la méthode de référence n'en est pas vraiment une et, souvent, les patients dont le diagnostic est indéterminé sont exclus de l'analyse, ce qui surestime la performance du biomarqueur [4,20]. C'est le biais dont l'effet est le plus important [4];
- un biais de sélection existe lorsque des patients non consécutifs ou non inclus au hasard participent à l'étude ;
- l'absence d'aveugle pour le biomarqueur constitue un biais qui en général surestime la performance diagnostique, bien que cet effet soit en fait relativement modeste [4];
- le biais de vérification survient lorsque seuls les patients qui ont bénéficié de la méthode de référence sont inclus, lorsque tous n'en n'ont pas bénéficié, ou lorsque différentes méthodes

de références sont utilisées [4]; le biais est particulièrement gênant lorsque la décision de pratiquer le test de référence est influencée par le résultat du biomarqueur testé;

- il peut y avoir des résultats ininterprétables; malheureusement ces observations sont rarement rapportées, souvent exclues de l'analyse, introduisant un biais; pour les biomarqueurs dont l'interprétation peut être subjective (c'est le cas de nombreux tests effectués au lit du patient), la variabilité inter-observateur peut avoir un impact important bien que souvent ni rapportée ni estimée.

- la performance diagnostique peut s'améliorer avec le temps (entraînement du biologiste ou du médecin, amélioration de la technique); la mesure de la troponine Ic est un excellent exemple de l'amélioration progressive de la technique de mesure ces dernières années, modifiant le seuil de normalité du biomarqueur.

Comme pour les essais cliniques, il peut y avoir un biais de publication, seules les études positives avec des résultats encourageants étant publiées. Ce biais est important pour les méta-analyses, et, en l'absence de registre pour les études diagnostiques, il est impossible d'en estimer l'impact.

Enfin, dans une étude diagnostique, la méthode de référence devrait être une méthode étalon quasi parfaite ("gold standard") mais dans de nombreuses situations cliniques cette référence idéale n'existe pas (exemple de l'insuffisance cardiaque ou rénale) ou n'est pas envisageable (autopsie) ou n'est pas disponible chez tous les patients. Dans de nombreuses situations cliniques, le biomarqueur est comparé au résultat d'un score au lieu d'un diagnostic précis : score de Framingham pour l'insuffisance cardiaque, critères d'infection et de SIRS pour le sepsis, score de RIFLE pour l'insuffisance rénale. Lorsqu'une méthode de référence imparfaite est utilisée, il convient de reconnaître que les résultats peuvent être biaisés [21]. Plusieurs options existent lorsque la méthode de référence parfaite n'existe pas ou ne peut être utilisée. On peut utiliser le consensus d'un groupe d'experts pour établir le diagnostic de référence qui doit avoir un accès à l'ensemble de l'information disponible, à l'exception de la valeur du biomarqueur testé. L'accord entre les experts doit être rapporté dans l'étude. Une autre possibilité est d'associer une probabilité du diagnostic (de 0 à 1) à chaque patient, établie subjectivement ou à l'aide d'un modèle de régression logistique intégrant plusieurs variables. Enfin, il est possible de transformer un problème diagnostique en un problème pronostic.

5. La démarche de recherche est-elle scientifique ?

Très souvent, les études diagnostiques sur un biomarqueur se contentent d'une « évaluation » de celui-ci, ce qui malheureusement s'apparente plus à un lancer de filets pour voir ce qu'il

est possible d'attraper que d'une véritable démarche scientifique. En tout état de cause, les principes fondamentaux de la démarche scientifique expérimentale doivent être suivis : établir une hypothèse quantifiable testable (« réfutable » au sens de Karl Popper) et la tester. Ainsi, toute étude devrait clairement expliquer au lecteur sur quel critère de jugement principal (et donc unique) elle compte s'appuyer, et décrire donc une hypothèse quantifiable précise : la sensibilité du biomarqueur X doit être supérieure à 0,75, l'aire sous la courbe ROC du biomarqueur Y doit être supérieure à celle du test utilisé habituellement, etc. En cas de comparaison avec une autre technique, il doit être également clairement établi s'il s'agit d'une étude de supériorité, d'équivalence, ou de non infériorité.

Bien que souvent passé sous silence [22], le problème de la puissance des études est majeur dans les études diagnostiques comme dans le reste de la recherche clinique [4]. Toute étude diagnostique devrait comporter un calcul a priori du nombre de patients à inclure, calcul qui découle très logiquement de l'hypothèse que l'on souhaite tester. Le calcul est rendu très facile par l'existence de nombreux logiciels adaptés. Une revue des procédés et des tables de calcul ont été proposées [23] et il existe également de nombreuses ressources de calcul disponible sur Internet. Il convient d'aller au-delà de la sensibilité et de la spécificité. Il est possible d'effectuer un calcul de nombre de patients en considérant la sensibilité avec une proportion donnée de faux positifs, l'aire sous la courbe ROC, et les indices de reclassification. L'objectif peut être aussi de comparer des seuils d'un biomarqueur ou de comparer plus de deux biomarqueurs. En fait, l'évaluation de biomarqueurs peut comprendre des analyses statistiques très différentes nécessitant des calculs de puissance différents. Bien que les techniques de calcul d'effectif ne soient pas toujours disponibles dans les logiciels disponibles sur le marché, des logiciels libres sont disponibles (R software, <http://cran.r-project.org/>) bien que souvent peu accessibles aux non initiés. L'aide d'un biostatisticien devient alors nécessaire.

7. Validation interne et validation externe

Lorsqu'un biomarqueur fait l'objet d'une étude, une question importante se pose à l'investigateur : quelle est la validité de l'étude. La validité externe passe impérativement par l'étude d'une cohorte différente de celle qui a permis l'observation des résultats. Ceci est rarement possible au sein d'une même étude [24] et nécessite en fait le plus souvent de nouvelles études effectuées par d'autres investigateurs. Une pratique encore répandue consiste à diviser la cohorte initiale en deux cohortes l'une permettant d'obtenir des résultats (cohorte dite de dérivation) et l'autre permettant de les valider (cohorte de validation). Cette pratique

est actuellement vivement critiquée sur le plan statistique, et avec raison, car les deux sous-populations sont par définition semblables car issues de la même population initiale.

En revanche, les techniques de rééchantillonnage permettent d'obtenir une validation interne des résultats particulièrement intéressante. La technique de bootstrap permet ainsi de générer des centaines ou des milliers de courbes ROC à partir d'une même population, permettant par exemple de mieux définir l'intervalle de confiance de l'aire sous la courbe ROC ou d'obtenir un intervalle de confiance du seuil diagnostique [5]. De plus en plus, les journaux prestigieux exigent l'utilisation des techniques de ré-échantillonnage afin de permettre au moins une validation interne des résultats et une donc une démonstration de leur robustesse (Fig. 1) [5]. De toute façon, à terme, une validation externe (autre population, autres conditions) est incontournable.

8. Nocivité des biomarqueurs

C'est un sujet rarement abordé dans la littérature et en pratique quotidienne, considérant de manière erronée que la prescription d'un biomarqueur ne peut avoir que des effets bénéfiques. Or il faut considérer que toute action en médecine est susceptible d'avoir des effets bénéfiques ou délétères (effets adverses). Il est aisé de comprendre que la prescription d'un biomarqueur a des effets délétères lorsque le diagnostic est erroné (faux positifs et faux négatifs). Toutefois, on sous-estime fréquemment les effets d'une prescription induite lorsque la question ne se posait pas vraiment et que la prescription a été systématique ou effectuée tout simplement par excès. Ainsi la prescription de D-dimères non justifiée (par exemple lorsque le résultat positif ne sera pas interprétable a priori) peut-elle induire une cascade d'événement dont la prescription de nouveaux examens (angioscanner) non dépourvus de risque pour le patient. C'est la principale raison pour laquelle les études d'intervention randomisée (avec et sans prescription du biomarqueur) sont aussi importantes. Malheureusement elles sont aussi rares.

8. Conclusion

En dehors d'un nombre très réduit de biomarqueurs (lactate, procalcitonine, BNP, troponine) qui ont fait l'objet de très nombreuses publications internationales de haut niveau y compris des études interventionnelles, il faut reconnaître que le niveau de preuve pour les nouveaux biomarqueurs reste encore souvent insuffisant. L'étude d'un biomarqueur est trop souvent effectuée avec une méthodologie générale et statistique médiocre qui limite considérablement la portée scientifique du message et sa reproductibilité. Parfois, c'est l'ensemble de la démarche de recherche qui n'est guère scientifique en l'absence d'une hypothèse clairement

identifiable, quantifiée, et testable. Cette constatation contraste avec un développement considérable de l'activité scientifique et industrielle explorant de nouveaux biomarqueurs cardiaques, rénaux, neurologiques, de l'inflammation ou du sepsis. Des recommandations ont été récemment publiées mais elles ne couvrent pas encore l'ensemble du domaine [5]. Deux raisons principales expliquent ce retard: premièrement, il y a un retard inévitable entre le développement des outils biostatistiques et leur application en recherche clinique. Deuxièmement, même en recherche biostatistiques, le domaine des études diagnostiques accuse un certain retard par rapport à d'autres domaines. Il y a donc un besoin urgent d'améliorer nos méthodes d'analyse des performances diagnostiques des biomarqueurs, particulièrement en ce qui concerne l'utilisation des courbes ROC, le choix des seuils, la définition de zone grise, le calcul du nombre de patients à inclure, et les techniques de validation interne. Les études sur les biomarqueurs peuvent présenter des limitations méthodologiques très importantes : les investigateurs doivent les prendre en compte dans la construction de leurs études, les rédacteurs en chef et les relecteurs des journaux quand ils les analysent et les acceptent pour publication, et les lecteurs quand ils en prennent connaissance.

Références

1. Baker M. In biomarker we trust? *Nature Biotechnol* 2005; 23: 297-304
2. Riou B. Troponin: Important in severe trauma and a first step in the biological marker revolution. *Anesthesiology* 2004; 101: 1259-60
3. Bossuyt PM, Reitsma JR, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Ann Intern Med* 2003; 138: W1-W12
4. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282: 1061-6
5. Ray P, Le Manach Y, Riou B, Houle T. Statistical evaluation of a biomarker. *Anesthesiology* 2010; 112: 1024-32
6. Riou B. Principales recommandations pour l'évaluation d'un biomarqueur. In : *Les biomarqueurs*, Ray P, Claessens YE Eds, Collection Référence en Médecine d'Urgence, Société Française de Médecine d'urgence, Paris, 2012.
7. Zweig MH, Campbell G. Receiver-operating characteristics (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993; 39: 561-77
8. Christ-Crain M, Jaccard-Stoltz D, Bingisser R, et al. Effect of procalcitonin-guided treatment on antibiotic use and outcome in lower respiratory tract infections: cluster-randomized, single-blinded intervention trial. *Lancet* 2004; 363: 600-7
9. Schneider HG, Lam L, Lokuge A, et al. B-type natriuretic peptide testing, clinical outcomes, and health services use in emergency department patients with dyspnea: a randomized trial. *Ann Intern Med* 2009; 150: 365-71
10. Nobre V, Harbarth S, Graf JD, Rohner P, Pugin J. Use of procalcitonin to shorten antibiotic treatment duration in septic patients: a randomized trial. *Am J Respir Crit Care Med* 2008; 177: 498-505
11. Marshall JC, Reinhardt K, for the International Sepsis Forum. Biomarkers of sepsis. *Crit Care Med* 2009; 37: 2290-8
12. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134: 663-94
13. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic tests research. Getting better but still not good. *JAMA* 1995; 274: 645-51
14. Obuchowski NA, Lieber ML, Wians FH. ROC curves in *Clinical Chemistry*: Uses, misuses, and possible solutions. *Clin Chem* 2004; 50: 1118-25
15. Saah AJ, Hoover DR. « Sensitivity » and « specificity » reconsidered : the meaning of the terms in analytical and diagnostic settings. *Ann Intern Med* 1997; 126: 91-4

16. Mower WR. Evaluating bias and variability in diagnostic test. *Ann Emerg Med* 1999; 33: 85-91
17. Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: An old concept in a new setting. *Am J Epidemiol* 2008 ; 168: 89-97
18. Trinquart L, Ray P, Riou B, Texeira A. Natriuretic peptide testing in EDs for managing acute dyspnea: a meta-analysis *Am J Emerg Med* 2011; 29 : 757-67
19. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987; 6: 411-23
20. Fischer JE, Bachmann LM, Jaeschke R. A reader's guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med* 2003; 29: 1043-51
21. Valenstein PN. Evaluating diagnostic tests with imperfect standards. *Am J Clin Pathol* 1990; 93: 252-8
22. Bachmann LM, Puhan MA, ter Riet G, Bossuyt PM. Sample sizes of studies on diagnostic accuracy: literature survey. *BMJ* 2006; 332: 1127-9
23. Flahaut A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005; 58: 859-62
24. Sartorius D, Le Manach Y, David JS, et al. Mechanism, Glasgow Coma Scale, Age, Arterial Pressure (MGAP): A new simple prehospital triage score to predict mortality in trauma patients. *Crit Care Med* 2010; 38: 831-7
25. Ray P, Arthaud M, Lefort Y, et al. Usefulness of B-type natriuretic peptide in elderly patients with acute dyspnea. *Intensive Care Med* 2004; 30: 2230-6

Tableau 1: Principales utilisations d'un biomarqueur. D'après Ray et al. [5].

Utilisation	Description	Exemples
Diagnostic d'une maladie	Permettre un diagnostic plus fiable, plus rapide, ou plus économique que les méthodes disponibles	La troponin Ic diagnostique l'infarctus du myocarde La procalcitonine diagnostique les infections bactériennes
Évaluation de la sévérité	Identifier un sous-groupe de patients avec une forme sévère de la maladie associée à un mauvais pronostic	La procalcitonine identifie le patient septique à mauvais pronostic La troponin Ic identifie les formes sévères de l'embolie pulmonaire
Évaluation du risque	Identifier un sous-groupe de patients ayant un pronostic différent lorsqu'ils sont exposés à une intervention	Un BNP élevé est associé à un mauvais pronostic après chirurgie non cardiaque
Prédiction de l'effet d'un médicament	Identifier une réponse pharmacologique chez un patient exposé à un médicament (efficacité, toxicité, pharmacocinétique)	Efficacité du clopidogrel
Monitoring	Évaluer la réponse à une intervention thérapeutique	La procalcitonine permet de guider l'arrêt d'une antibiothérapie

Table 2: La check liste STARD pour les études diagnostiques. D'après Bossuyt et al. [3].

Section	Item	Description
Titre, résumé, mots clés	1	Identifier l'article comme une étude diagnostique (termes MeSH recommandée "sensibilité et spécificité"
Introduction	2	Énoncer la question posée ou l'objectif de la recherche en terme d'évaluation de la performance diagnostique ou de comparaison de celle-ci entre tests ou entre groupes de patients
Méthodes, participants	3	Décrire la population étudiée: critères d'inclusion et d'exclusion, caractéristiques et localisation des structures où les données ont été collectées
	4	Décrire le recrutement des patients: selon un symptôme, selon les résultats de tests, selon que les patients ont bénéficié du test de référence ou du biomarqueur testé
	5	Décrire l'échantillonnage des patients : patients consécutifs selon les critères des items 3 et 4 ? Sinon, comment les patients ont-ils été sélectionnés ?
	6	Décrire le recueil des données: le recueil a-t-il été planifié avant celui de la méthode de référence ou du biomarqueur testé (prospectif) ou après (rétrospectif) ?
Méthode de référence	7	Décrire la méthode de référence et sa justification
	8	Décrire les techniques de mesures spécifiques, incluant quand et comment elles ont été effectuées, et citer les références pour la méthode de référence et la méthode testée, ou les deux
	9	Décrire les unités, les seuils choisis, ou les catégories, la méthode de référence et la méthode testée
	10	Décrire le nombre, l'entraînement, et l'expertise des personnes effectuant la méthode de référence et la méthode testée
	11	Est-ce que les investigateurs évaluant la méthode de référence et la méthode testée étaient aveugles vis-à-vis des autres données ? Décrire les informations cliniques dont ils disposaient.
Méthodes statistiques	12	Décrire les méthodes pour évaluer ou comparer la performance diagnostique et les méthodes statistiques utilisées pour estimer l'incertitude (intervalle de confiance à 95%)
Résultats, participants	13	Décrire les méthodes utilisées pour quantifier la reproductibilité du test
	14	Décrire la période de l'étude, en incluant le début et la fin de la période de recrutement.
	15	Décrire les caractéristiques démographiques et cliniques de la population (âge, sexe, symptômes, comorbidité, traitements, et centres participants)
	16	Décrire combien de participants répondant aux critères d'inclusion ont ou n'ont pas bénéficié de la méthode de référence et de la méthode testée, ou des deux ; expliquer pourquoi certains n'en ont pas bénéficié (« flow diagram »).
Résultats du test	17	Décrire l'intervalle de temps entre la méthode de référence et la méthode testée, et les traitements administrés entre.
	18	Décrire la répartition des formes de sévérité de la maladie (critères définis) chez les patients malades et les autres diagnostics présents chez les patients non malades
	19	Décrire dans un tableau de contingence les résultats de la méthode testée (incluant les résultats indéterminés et manquants) en fonction des résultats de la méthode de référence ; pour les variables continues, décrire la distribution de la méthode testée en fonction des résultats de la méthode de référence.
	20	Décrire tous les effets indésirables liés à la pratique de la méthode de référence et de la

		méthode testée.
Estimation	21	Estimer la performance diagnostique et exprimer la variabilité statistique du résultat obtenu (intervalle de confiance à 95%)
	22	Décrire les résultats indéterminés, les valeurs manquantes, et comment les valeurs extrêmes anormales ont été considérées.
	23	Estimer la variabilité entre investigateurs, entre centres, éventuellement entre sous-groupes.
	24	Estimer la reproductibilité du test
Discussion	25	Discuter de l'applicabilité clinique des résultats de l'étude