

Les scores et le jugement clinique

B. Housset

Les travaux publiés proposant des règles de prédiction clinique sont nombreux, qu'il s'agisse de l'estimation d'une probabilité diagnostique ou pronostique. Ces règles de prédiction peuvent se présenter sous la forme d'algorithmes (« si... alors... »), ou d'équations de régression dont l'expression peut être un score qui utilise la somme de variables sélectionnées et pondérées. Comme le rappellent A. Perrier et T. Peneger [1] dans le cinquième article de la série « médecine factuelle » de la *Revue des Maladies Respiratoires* [2-6], les scores cliniques doivent répondre à des critères méthodologiques stricts d'abord de dérivation puis de validation. Toutefois, il est surprenant de voir que ces scores sont peu utilisés en pratique [7]. Si la méthodologie utilisée pour établir un score est pratiquement toujours précisément décrite dans la littérature, l'analyse d'impact fait beaucoup plus rarement l'objet d'une publication. Ainsi sur 30 règles de prédiction clinique publiées de 1991 à 1994, seuls 3 % des publications précisaient l'impact clinique de ces scores [8].

Pourquoi ces règles de prédiction clinique ne sont-elles que peu ou pas utilisées en pratique clinique ? Plusieurs raisons peuvent être avancées. Tout d'abord la crédibilité clinique des modèles proposés suppose quelques pré-requis.

Premièrement, les signes cliniques pertinents doivent être intégrés aux scores. Que penserait-on d'un score estimant la probabilité de malignité de nodules pulmonaires et ne prenant pas en compte la notion de tabagisme ?

Deuxièmement, les variables utilisées pour calculer le score doivent être faciles à recueillir à un moindre coût. Par exemple, pour estimer la probabilité d'embolie pulmonaire à partir du score discuté dans l'article de Perrier et Perneger [1] et publié par Wicki et coll. [9], il est nécessaire de connaître les résultats des gaz du sang artériel pour établir la probabilité d'embolie pulmonaire. Ceci limite d'emblée les possibilités d'utilisation de ce test en médecine générale. De plus, le recueil des signes doit être fiable et reproductible d'un médecin à un autre. L'estimation du score de Glasgow, destiné à évaluer la profondeur d'un coma, est, par exemple, fortement dépendante de l'entraînement du clinicien [10]. Plus récemment, à propos d'un score clinique de diagnostic d'angine d'origine

Service de Pneumologie et de Pathologie professionnelle, Centre Hospitalier Intercommunal de Créteil, France.

Tirés à part : B. Housset, Service de Pneumologie et de Pathologie professionnelle, Centre Hospitalier Intercommunal de Créteil, 40, avenue de Verdun, 94010 Créteil, France.
Bruno.Housset@chicreteil.fr

Réception version princeps à la Revue : 22.01.2003.

Acceptation définitive : 29.01.2003.

Publié par anticipation sur www.splf.org/rmr/ le 18.04.2003.

streptococcique, des auteurs néerlandais ont montré le manque de concordance entre cliniciens dans le recueil des signes [11]. Il est intéressant de souligner à ce propos que les scores cliniques renvoient à la sémiologie, pierre angulaire de la médecine clinique. C'est un de leurs intérêts majeurs que d'amener à poser les bonnes questions et à rechercher avec rigueur les signes importants pour le diagnostic ou le pronostic.

Troisièmement, les variables utilisées dans un score ne doivent pas faire appel, dans la mesure du possible, à des notions de seuil pour des variables continues. Par exemple, Wicki et coll. [8] stratifient la PaO₂ en 4 classes et la PaCO₂ en 2 classes. Il est difficile de croire qu'une PaO₂ de 59 ou de 61 mm Hg ait un poids différent dans l'estimation de la probabilité diagnostique.

Quatrièmement, le score doit être facile à calculer. Une fois refermé ce numéro après une lecture attentive de l'article de Perrier et coll., quel clinicien aura la capacité mnémotique lui permettant d'utiliser le score de Wicki sans l'aide minimale d'un papier et d'un crayon ? Encore s'agit-il d'un score dont le nombre d'items est relativement limité. L'application du score de Fine pour l'appréciation pronostique d'une pneumonie fait appel à un plus grand nombre de variables, pour chacune pondérée [12]. L'ordinateur de bureau et la confection d'une macro-commande sur un tableur de type Microsoft Excel® fournissent une solution mais c'est probablement le développement des ordinateurs de poche, ou assistants électroniques personnels (de type Palm® ou Pocket PC®) qui offrent la possibilité d'un calcul du score en consultation, au lit du patient [13, 14]. Citons le programme Medcalc®, d'origine genevoise, disponible gratuitement en français, anglais et espagnol [15]. Ce programme de calcul fournit aux cliniciens de nombreuses équations et scores cliniques. Construire ses propres scores à partir de données publiées est devenu d'une grande facilité sur ces ordinateurs de poche. En effet, des tableurs et des bases de données aisément configurables sont maintenant disponibles sur ce type d'appareil. Nous renvoyons le lecteur intéressé vers le numéro 53 d'*Info Respiration* [16, 17] ou vers le site internet de la SPLF (www.splf.org) où l'auteur de ces lignes propose le téléchargement de quelques scores cliniques potentiellement utiles aux pneumologues.

Cinquièmement, la structure du modèle doit être aisément compréhensible pour le clinicien. Fournir une équation de régression parlera moins qu'une série d'items cliniques associée à un coefficient de pondération. De plus, l'évolution des modèles vers la complexité, tels les modèles de réseaux neuronaux¹, risque de constituer par leur opacité, une cause de

¹ Un réseau neuronal associe de nombreux processeurs élémentaires simples, analogues à des neurones, dont le comportement collectif remplit une fonction souhaitée par l'utilisateur. Celui-ci doit procéder à l'apprentissage du réseau, c'est à dire le modifier progressivement en lui soumettant des exemples du comportement attendu. Une fois l'apprentissage du réseau achevé, ce dernier se comporte comme une « boîte noire » à laquelle on peut soumettre de nouvelles données, et qui est censé fournir les réponses appropriées.

résistance à l'appropriation des règles de prédiction clinique. Un score clinique doit être précis en assurant un faible taux de faux positifs (prédiction d'un événement qui ne se réalise pas) et de faux négatifs (absence de prédiction d'un événement qui se réalise). On voit ici combien un score clinique peut se comparer à un test diagnostique et la nécessité d'une validation externe aussi extensive que possible sur des populations variées. C'est ainsi qu'ont été proposés des niveaux de preuve pour les scores cliniques [16]. Cela fleurit bon la médecine factuelle ! La validation sur des populations distinctes dans l'espace et dans le temps de la population de dérivation est une assurance sur la possible généralisation du score clinique. C'est souligner ici la nécessaire rigueur de la description des items recueillis et des caractéristiques de la population étudiée.

Enfin le médecin doit être convaincu de l'efficacité clinique de la règle de prédiction. Le test doit répondre en pratique à ce pourquoi il a été conçu, à savoir aider le clinicien à prendre la meilleure décision. Le score se doit donc d'être « meilleur » que le raisonnement du clinicien, ultime obstacle à l'utilisation d'une règle de prédiction [18]. En effet, le clinicien intègre dans son raisonnement l'expérience clinique qui comporte de nombreux biais maintenant bien identifiés, qu'il s'agisse de biais de mémorisation ou de biais affectifs. C'est une étape parfois psychologiquement difficile que de passer de l'implicite d'une expérience clinique limitée au caractère explicite, neutre et cartésien, d'un score clinique établi et validé sur une vaste population. L'explicite est une dimension essentielle de la médecine fondée sur des preuves. Un des avantages et non des moindres de l'explicite est de pouvoir être analysé, critiqué et éventuellement modulé en fonction d'un contexte clinique particulier... ce qui donne une place essentielle à la rigueur et à l'objectivité du jugement clinique ! Ceci peut même conduire le clinicien jusqu'à rejeter sur une argumentation explicite le résultat d'un score dont l'application et/ou le résultat ne lui semblent pas en cohérence avec une situation clinique précise.

Dans la perspective du développement d'une médecine factuelle, « fondée sur des preuves », la mise en œuvre de scores cliniques validés et efficaces est un enjeu important. Merci à Arnaud Perrier et Thomas Peneger d'appuyer cette démarche par leur contribution à ce numéro.

Références

- 1 Perrier A, Perneger T : Les règles de prédiction : un guide de lecture. *Rev Mal Respir* 2003 ; 20 : 249-56.
- 2 Perrier A, Similowski T : Une série sur la médecine factuelle dans la *Revue des Maladies Respiratoires* : mais pour quoi faire ? *Rev Mal Respir* 2002 ; 19 : 395-8.
- 3 Perrier A : L'essai randomisé-contrôlé : un guide de lecture. *Rev Mal Respir* 2002 ; 19 : 505-14.
- 4 Sculier JP : Méta-Analyse : un guide de lecture. *Rev Mal Respir* 2002 ; 19 : 633-7.

- 5 Nendaz MR, Perrier A : Etude de validation d'un test diagnostique : un guide de lecture. *Rev Mal Respir* 2002 ; 19 : 767-77.
 - 6 Vergnenègre A : L'analyse coût-efficacité : un guide de lecture. *Rev Mal Respir* 2003 ; 20 : 116-25.
 - 7 Wyatt JC, Altman DG : Commentary: Prognostic models: clinically useful or quickly forgotten? *BMJ* 1995 ; 311 : 1539-41.
 - 8 Laupacis A, Sekar N, Stiell IG : Clinical prediction rules. A review and suggested modifications of methodological standards. *Jama* 1997 ; 277 : 488-94.
 - 9 Wicki J, Perneger TV, Junod AF, Bounameaux H, Perrier A : Assessing clinical probability of pulmonary embolism in the emergency ward: a simple score. *Arch Intern Med* 2001 ; 161 : 92-7.
 - 10 Rowley G, Fielding K : Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *Lancet* 1991 ; 337 : 535-8.
 - 11 Donner-Banzhoff N, Beck C, Meyer F, Werner JA, Baum E : Clinical findings in patients presenting with sore throat. A study on inter-observer reliability. *Fam Pract* 2002 ; 19 : 466-8.
 - 12 Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, Coley CM, Marrie TJ, Kapoor WN : A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997 ; 336 : 243-50.
 - 13 Nesbitt TS, Jerant A, Balsbaugh T : Equipping primary care physicians for the digital age. The Internet, online education, handheld computers, and telemedicine. *West J Med* 2002 ; 176 : 116-20.
 - 14 Larkin M : Handheld use increasing for e-learning and clinical decision making. *Lancet* 2003 ; 361 : 93.
 - 15 Tschopp M : Medcalc <http://medcalc.med-ia.net/>.
 - 16 Rouzaud P : Peut-on prédire la thrombose veineuse et l'embolie pulmonaire par la clinique ? *Info-Respiration* 2003 ; n° 53 : 14-5.
 - 17 Housset B : « PDA » et scores cliniques. *Info-Respiration* 2003 ; n° 53 : 16.
 - 18 Junod A : Des scores cliniques ou de l'urgence laborieuse de l'explique dans la décision médicale. *Med Hyg* 2001 ; 59 : 2099-100.
-