

# Evidence-based Medicine: comment juger une étude sur un test diagnostique?

## Evidence-based Medicine: How to read an article about a diagnostic test

U. Glenck<sup>a</sup>, D. Pewsner<sup>b</sup>, H. C. Bucher<sup>c</sup>

### Introduction

Les progrès technologiques ont permis de réaliser d'innombrables nouvelles possibilités de diagnostic et d'en améliorer d'autres au cours de ces trente dernières années. Elles sont même devenues si nombreuses que les médecins courent le risque de faire de personnes en parfaite santé des malades, en pratiquant des tests sans discernement. Et cela est particulièrement vrai lorsque qu'il s'agit de tests de dépistage dans une population saine selon des critères cliniques. En plus du travail médical compétent, avec la santé de chaque patient individuellement et de groupes entiers à l'esprit, la prise de conscience toujours aiguë de la limitation des ressources économiques exige d'utiliser ces possibilités diagnostiques de manière aussi rationnelle que ciblée.

Au vu de cette évolution, l'information sur la performance des tests diagnostiques pour les médecins hospitaliers a une importance capitale. Mais nombreux sont ces cliniciens à être trop peu informés sur les questions et critères essentiels d'appréciation de la qualité de ces tests, ou qui n'ont pas le temps de lire cette information, ni de l'évaluer de manière critique. Et comme les résultats de la recherche clinique sont actuellement à la portée des collègues installés en périphérie, ce travail montrera comment juger les études cliniques sur les tests diagnostiques, et comment en tirer les conclusions correctes pour leur pratique quotidienne.

### Scénario de la pratique quotidienne

Vous recevez en consultation une femme de 27 ans inquiète à propos d'une de ses jambes qui a enflé. Ce symptôme n'est apparu qu'il y a 36 heures, et pour la première fois. Elle ne fume

pas, elle est mère d'un petit garçon de deux ans et prend un inhibiteur de l'ovulation. Son anamnèse familiale est négative pour les maladies thromboemboliques. L'examen montre un léger œdème malléolaire ne prenant pas le godet, aucune douleur particulière à la pression des mollets et une différence de circonférence de 1,5 cm par rapport à sa jambe saine. Mis à part quelques veinectasies depuis sa grossesse, elle ne présente pas de varices.

### Problème de diagnostic

Le risque que cette patiente présente une thrombose veineuse profonde (TVP) est franchement minime. Mais vous vous souvenez d'un cas semblable, où vous avez raté une thrombose. Cette fois vous voulez être plus sûr avant de laisser filer cette patiente sans anticoagulation. Une phlébographie en urgence vous semble exagérée. Alors une ultrasonographie duplex? Vous avez reçu récemment un kit de test pour doser les D-dimères (SimpliRed®). Et comme ce test est négatif, vous la laissez partir sans diagnostic, mais après avoir exclu une TVP. Dans la soirée, vous parcourez la littérature que vous aviez préparée, et tombez dans «Ars Medici» [1] sur un article dans lequel les auteurs prétendent que le dosage des D-dimères n'est pas indiqué pour exclure une thrombose. Avec une sensibilité de 80%, vous passez à côté d'une TVP sur cinq avec un résultat négatif!

### Question clinique

Vous hésitez, et vous vous posez la question suivante: quelle est la fiabilité du dosage des D-dimères (SimpliRed®) pour exclure une TVP avec suspicion clinique faible ou moyennement forte?

<sup>a</sup> Médecine générale, Ottenbach

<sup>b</sup> Médecine interne, Bern

<sup>c</sup> Policlinique médicale universitaire, Hôpital cantonal de Bâle

Correspondance:

Dr Heiner C. Bucher, PD  
Policlinique médicale universitaire  
Hôpital cantonal de Bâle  
CH-4031 Bâle

E-mail: [hbucher@uhbs.ch](mailto:hbucher@uhbs.ch)

## Recherche de littérature

Après avoir soupé, vous allumez votre ordinateur pour faire une recherche dans la littérature. Par Internet, vous avez accès à Medline (<http://www.hin.ch>). Et comme vous voulez en savoir plus sur le test que vous avez effectué, vous entrez «D-dimers», «simplired» et «thrombosis», dans la recherche et recevez 25 travaux sur ce thème.

Un bref passage sur les titres et résumés de ces travaux vous apprend qu'ils donnent des conclusions contradictoires sur l'utilité du dosage des D-dimères SimpliRed® dans l'exclusion des TVP [2, 3]. A quels travaux vous fier? Vous vous souvenez d'avoir lu dans «Evidence-Based Medicine» [4] une discussion de deux travaux ayant examiné l'utilité du test rapide des D-dimères [5, 6]. Cette revue présente et commente des travaux originaux satisfaisant certains critères de qualité. Dans cet article, nous allons regarder d'un peu plus près l'un de ces travaux, provenant de l'Hôpital cantonal de Bâle. Vous pouvez charger le travail original sur le site du Journal of Vascular Surgery (<http://www.harcourthealth.com>), après l'avoir trouvé avec votre browser web.

## Vérifier la preuve («critical appraisal»)

A quels résultats d'études pouvons-nous nous fier? Sackett et al. ont développé des critères de

«critical appraisal» pour les études sur les tests diagnostiques [7]. Il faut pouvoir répondre à trois grandes questions:

- La preuve de la précision du test diagnostique est-elle valide, et a-t-elle été correctement vérifiée méthodologiquement?
- Ce test permet-il de distinguer les patients qui ont la maladie recherchée de ceux qui ne l'ont pas?
- Puis-je utiliser ce test pour mon patient? M'est-il utile pour décider de demander d'autres examens diagnostiques ou de prescrire un traitement?

## La preuve de la précision du test diagnostique est-elle valide, et a-t-elle été correctement vérifiée méthodologiquement?

Cette première question concerne la qualité méthodologique de l'étude. Si un test n'a pas été examiné selon une méthodologie suffisante, vous pouvez abandonner la lecture. La qualité de la méthode se vérifie en 4 points (Tableau 1).

– *Y a-t-il une comparaison en aveugle, indépendante, avec un test standard («gold standard»)?*

C'est la méthode utilisée dans l'étude qui permet de savoir si ses résultats sont dignes de confiance. Nous devons savoir dans quelle mesure le test, et les résultats du test, reflètent des situations bien réelles. Nous devons nous assurer que le test à évaluer a été comparé à un test de référence valable, ou à un test standard. Le test de référence doit être celui qui a fait la preuve de la plus grande fiabilité («gold standard»), par exemple une biopsie, une autopsie ou un suivi à long terme. Si le test de référence semble valable, il faut encore vérifier que ses résultats et ceux du test à examiner ont été évalués de manière **indépendante**. C'est-à-dire que ceux qui lisent les résultats du test doivent le faire **en aveugle** («blinded assessment»), soit sans connaître les résultats du gold standard, et inversement. Plus grande est la probabilité que l'interprétation du test soit faussée par la connaissance du résultat du test de référence, ce qui est surtout le cas de l'imagerie diagnostique, plus la méthode à l'aveugle est importante.

Dans notre exemple, cette condition est remplie. Les interprètes du test SimpliRed® ne savaient rien des résultats de la sonographie duplex, et inversement. Le gold standard utilisé a été la sonographie duplex, malgré le fait que ce soit habituellement la phlébographie. Les auteurs justifient leur position en disant que dans leur service, la sensibilité (94–97%) et la spéci-

**Tableau 1.**  
Critères d'appréciation de la méthode d'une étude sur un test diagnostique [7].

### La preuve de la précision du test diagnostique est-elle valide, et a-t-elle été correctement vérifiée méthodologiquement?

- Y a-t-il une comparaison indépendante, en aveugle, avec un test standard («gold standard»)?
- Cette étude a-t-elle incorporé un spectre suffisamment large de patients pour lesquels ce test est utilisé en pratique clinique?
- Les résultats du test à évaluer ont-ils influencé la décision de recourir à un test standard?
- La méthode du test est-elle décrite suffisamment en détail pour en permettre la répétition? Ce test a-t-il été examiné auprès d'un autre collectif, indépendamment de la présente étude?

### Ce test permet-il de distinguer les patients qui ont la maladie recherchée de ceux qui l'ont pas?

### Puis-je utiliser ce test pour mon patient? M'est-il utile pour décider de demander d'autres examens diagnostiques ou de prescrire un traitement?

- La reproductibilité et l'interprétation des résultats du test sont-elles données dans mon contexte clinique?
- Est-il possible d'estimer une probabilité prétest cliniquement raisonnable pour la maladie cible de mon patient?
- La probabilité posttest est-elle dans un domaine qui va modifier la prise en charge de mon patient?

ficité (91–96%) de cette technique sont supérieures à celles de la phlébographie. Et comme le gold standard doit être utilisé chez tous les patients, la méthode moins invasive de la sonographie duplex est parfaitement défendable. Méthodologiquement cependant, la fiabilité de la sonographie duplex est plus faible que celle de la phlébographie dans le diagnostic de la TVP. Dans notre exemple concret, cela peut avoir embelli la fiabilité réelle du test des D-dimères. Le fait de ne pas avoir recouru au gold standard a peut-être fait manquer des thromboses discrètes, voulant que la sensibilité et la spécificité de ce test pourraient avoir été surestimées.

*– L'étude a-t-elle incorporé des patients correspondant à ceux chez lesquels le test est utilisé en pratique?*

Un test diagnostique n'est utile que dans la mesure où il permet de distinguer parmi les caractéristiques cibles recherchées lesquelles ne seraient sans cela pas identifiables. Pratiquement n'importe quel test diagnostique permet de faire la distinction entre gravement malade et en bonne santé. Le véritable intérêt diagnostique d'un examen n'est donc illustré que dans une étude ayant incorporé des patients présentant la maladie cible voulue à un degré plus ou moins marqué. Ce qui correspond à notre expérience, car en pratique, les tableaux cliniques ne se présentent que rarement sous leur forme typique, comme dans les livres. Chez les patients ayant un carcinome colorectal par exemple, l'antigène carcino-embryonnaire (CEA) est nettement plus élevé que chez des personnes normales [8]. Ces résultats ont permis de supposer que ce dosage pourrait être indiqué comme test de dépistage du cancer colorectal. Les études effectuées sur un grand collectif de patients ayant un tel cancer à un stade peu avancé, d'autres tumeurs ou d'autres maladies de base, ont montré au contraire que le CEA donne souvent des résultats faux positifs. Le CEA a donc été abandonné comme test de dépistage.

Dans notre exemple, tous les patients ayant une suspicion de TVP et adressés à la Division d'Angiologie ont été examinés. Les critères d'exclusion sont clairement définis et les patients provenaient pour moitié de l'ambulatoire. Ces patients, recrutés dans les secteurs ambulatoire et hospitalier, ont eu des prévalences identiques de TVP proximale et distale, soit environ 5 et 15% respectivement. Le spectre de patients est donc vaste et représentatif de la pratique ambulatoire. Mais nous ne savons pas si un dosage des D-dimères avait déjà été effectué chez une partie des patients examinés avant leur admission, et donc si une sélection a été réalisée pouvant artificiellement améliorer la sensibilité de ce test.

*– Les résultats du test à évaluer ont-ils influencé la décision d'utiliser le test standard?*

Les propriétés d'un test à évaluer sauf faussées si ses résultats influencent la décision de faire subir au patient le test de référence ou non. Cette possibilité d'erreur est citée dans la littérature comme «verification bias». Voici un exemple pour illustrer cela. Dans une étude d'évaluation de la scintigraphie par ventilation-perfusion pour le diagnostic de l'embolie pulmonaire aiguë, les patients peu suspects d'avoir une embolie à la scintigraphie ont eu moins souvent une angiographie (69%) que ceux dont la scintigraphie était hautement évocatrice (92%) [9]. Il est bien compréhensible que les cliniciens ont moins tendance à exposer le patient à d'autres examens invasifs lorsque la probabilité d'obtenir davantage d'informations est très faible. Les auteurs de cette étude ont toutefois tenté de corriger ce biais en suivant sur une année les patients dont la scintigraphie était normale quant à la survenue d'embolies pulmonaires.

Dans cette étude, tous les patients ont passé par la sonographie duplex, «examen gold standard» (mais le véritable gold standard est la phlébographie). Les auteurs ont malheureusement négligé de s'assurer qu'ils n'avaient pas passé à côté d'une thrombose en suivant ces patients, ou en pratiquant un second examen chez les patients négatifs, ou chez ceux dont la probabilité de thrombose était très faible.

*– La méthodologie du test a-t-elle été décrite assez en détail pour en permettre la reproductibilité?*

*– Le même test a-t-il été effectué auprès d'un autre collectif, indépendamment de l'étude en question?*

Les études sur les tests diagnostiques doivent décrire très précisément la méthode, pour en permettre la reproductibilité et pouvoir vérifier que ce test soit utilisable en pratique clinique courante. Il faut que soient cités les indications importantes sur la technique, les méthodes de laboratoire ou le type d'appareils utilisés. L'évaluation d'un test dans un autre collectif de patients permet de donner des renseignements importants sur sa reproductibilité.

Les auteurs ont décrit en détail les étapes du test, pour qu'il soit reproductible, mais ils ont renoncé à l'appliquer à un autre collectif. Une deuxième étude de Londres [6] sur la même question, discutée dans le même numéro d'«Evidence-Based Medicine», arrive tout de même à des résultats comparables.

La conclusion est que cette étude, malgré certaines carences méthodologiques, permet de

dire que le dosage des D-dimères est suffisamment fiable pour exclure une thrombose veineuse profonde.

### Ce test permet-il de distinguer les patients qui ont la maladie recherchée de ceux qui l'ont pas?

Lorsque nous utilisons un test diagnostique, nous partons d'une suspicion clinique, soit d'une certaine probabilité (probabilité prétest) que la maladie cible soit présente. Le test que nous employons doit confirmer ou exclure cette maladie avec la plus grande certitude possible. Un résultat positif doit donner la probabilité posttest la plus grande, et un négatif la plus faible possible de la présence de la maladie suspectée. La qualité la plus importante du test s'exprime par sa sensibilité et sa spécificité. La «likelihood ratio» est un instrument moderne, plus valable que la sensibilité et la spécificité, et plus facile à concevoir dans l'appréciation des qualités d'un test. La likelihood ratio [7] décrit la performance d'un test, indépendamment de la probabilité prétest. Elle englobe les qualités du test, sa sensibilité et sa spécificité, dans un même chiffre et c'est un paramètre objectif

de la performance du test. Elle se définit comme le rapport des probabilités d'un résultat d'un certain test chez des malades et chez des personnes en bonne santé. La likelihood ratio pour un certain test positif (LR+) dit combien de fois un résultat positif du test est plus probable chez un malade que chez quelqu'un en bonne santé. La likelihood ratio pour un résultat négatif (LR-) dit combien de fois un résultat négatif du test est plus probable chez un véritable malade que chez quelqu'un en bonne santé.

La likelihood ratio pour un résultat positif est le rapport de la fréquence des résultats vrais positifs d'un test sur celle de ses faux positifs (tableau 2). La fréquence des vrais positifs, mesurée chez tous les patients ayant la maladie cible voulue, correspond à la sensibilité du test ( $a / a + c$ ). La fréquence des faux positifs ( $b / b + d$ ) mesurée chez toutes les personnes en bonne santé correspond à la valeur (1-spécificité). La LR+ est donc égale à sensibilité/1-spécificité.

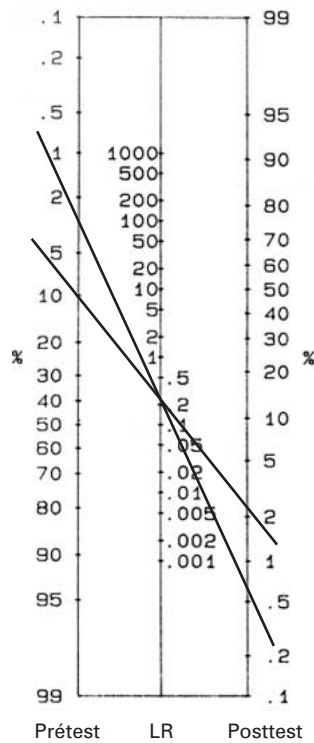
Le quotient de probabilité d'un résultat négatif (LR-) se définit comme le quotient de la fréquence des résultats faux négatifs sur celui des vrais négatifs (Tableau 2). La fréquence des faux négatifs ( $c / a + c$ ) correspond à la valeur (1-sensibilité), et celle des vrais négatifs à la spécificité d'un test ( $d / b + d$ ).

L'un des grands avantages de la likelihood ratio

**Tableau 2.**  
**Comparaison du test des D-dimères au test standard (sonographie duplex) pour le diagnostic d'une thrombose veineuse profonde (TVP) selon [7].**

Résultat du test à évaluer	Test de référence (sonographie duplex)		
	TVP présente	TVP absente	
D-dimères positifs	<b>a</b> Vrais positifs: <b>72</b>	<b>b</b> Faux positifs: <b>135</b>	Tous les tests positifs: <b>a + b: 207</b>
D-dimères négatifs	<b>c</b> Faux négatifs: <b>11</b>	<b>d</b> Vrais négatifs: <b>180</b>	Tous les tests négatifs: <b>c + d: 191</b>
	Tous les patients: <b>a + c: 83</b>	Tous les témoins <b>b + d: 315</b>	Toutes les personnes testées <b>a + b + c + d: 398</b>
<b>Sensibilité</b>	Fréquence de vrais positifs		
Nombre des malades ayant un test positif	$a / (a + c)$		$(72 / 83) \times 100\% = 86,7\%$
<b>Spécificité</b>	Fréquence de vrais négatifs		
Nombre des témoins ayant un test négatif	$d / (b + d)$		$(180 / 315) \times 100\% = 57,1\%$
<b>Likelihood ratio pour un test positif</b>	Fréquence des vrais positifs / Fréquence des faux positifs	Sensibilité / (1-spécificité)	$86,7\% / (100\% - 57,1\%) = 2,02$
<b>Likelihood ratio pour un test négatif</b>	Fréquence des faux négatifs / Fréquence des vrais négatifs	(1-sensibilité) / spécificité	$(100\% - 86,7\%) / 57,1\% = 0,23$
<b>Valeur prédictive positive (PPV)</b>	Nombre des tests vrais positifs sur tous les positifs		$(72 / 207) \times 100\% = 34,8\%$
<b>Valeur prédictive négative (NPV)</b>	Nombre des tests vrais négatifs sur tous les négatifs		$(180 / 191) \times 100\% = 94,2\%$
<b>Prévalence</b>	Probabilité prétest		$(83 / 398) \times 100\% = 20,8\%$

**Figure 1.**  
Nomogramme de Fagan [10].  
Estimation de la probabilité posttest sur la base de la probabilité prétest et de la likelihood ratio (LR).



est qu'avec la probabilité prétest, elle permet de calculer immédiatement la probabilité posttest d'un résultat de test. La probabilité prétest est également la prévalence d'une maladie cible dans une population donnée ou un groupe de patients donné. Le passage de la probabilité prétest à la posttest se fait le plus facilement du monde avec le nomogramme de Fagan [10] (Figure 1). Une autre source donne les calculs exacts [7, 11]. Les likelihood ratios ont des avantages importants sur les concepts usuels de sensibilité et de spécificité, du fait que pour un test donné, la probabilité posttest peut se calculer sur la base de plusieurs valeurs limites.

**Tableau 3.**  
**Dépendance de la sensibilité, de la spécificité et de la likelihood ratio des valeurs limites (cut-offs) à l'exemple de la sous-délevelation ST à l'ECG d'effort pour le diagnostic de la cardiopathie coronaire [15].**

Sous-délevelation ST en mm	Sensibilité (%)	Spécificité (%)	Likelihood ratio
< 0,5			0,23
0,5-0,99	86	77	0,92
1,0-1,49	65	89	2,1
1,5-1,99	42	98	4,2
2,0-2,49	33	99	11
> 2,5	20	>99	39

Des résultats de test peuvent donc être interprétés de manière plus différenciée avec les likelihood ratios, avec plusieurs valeurs limites (Tableau 3). Si par exemple le segment ST d'un patient est sous-délevelé de moins de 0,5 mm à l'ECG d'effort alors que la probabilité prétest était de 40%, la probabilité d'une cardiopathie coronaire chute du facteur 0,23 à 13%, et il n'y a donc pas d'indication à aller plus loin dans les examens. Dans une situation semblable, une sous-délevelation ST de 1,5 mm augmente la probabilité de 40 à 88%, indication évidente à l'angiographie coronaire.

Des likelihood ratios de l'ordre de >10 pour un test positif ou <0,1 pour un négatif donnent généralement des variations importantes de la probabilité prétest, permettant souvent (en fonction de la probabilité prétest) de confirmer ou d'exclure une maladie cible. Des likelihood ratios de 5 à 10 et de 0,1 à 0,2 resp. donnent des variations moyennes de la probabilité prétest. Et des quotients de probabilité de 2 à 5 et de 0,5 à 0,2 resp. donnent des variations minimes, mais parfois importantes de la probabilité prétest. Des likelihood ratios de 1 à 2 et de 0,5 à 1 ne font varier la probabilité prétest que de manière à peine significative cliniquement. Notre exemple montre que la likelihood ratio d'un résultat de test positif (LR+) pour les D-dimères est égale à 2,0 (Tableau 2). Ce résultat dit donc qu'une telle constellation est 2 fois plus fréquente chez un patient qui a effectivement une thrombose veineuse profonde (TVP) que chez un autre qui n'en a pas. Si la likelihood ratio est très proche de 1 (ce qui signifie probabilité prétest = probabilité posttest), le dosage des D-dimères n'a aucune importance pour confirmer le diagnostic de TVP. Par contre, la likelihood ratio pour un test D-dimères négatif est de 0,23, soit dans un domaine dans lequel une TVP peut être exclue avec suffisamment de certitude, surtout si la probabilité prétest est faible. Ce qui correspond à la sensibilité élevée de ce dosage, soit 87%. (Pour un test dont la sensibilité est élevée, si le résultat du test est négatif, la maladie cible recherchée peut être exclue avec une grande probabilité.)

**Puis-je utiliser ce test pour mon patient? M'est-il utile pour décider de demander d'autres examens diagnostiques ou de prescrire un traitement?**

Si les aspects méthodologiques d'une étude d'évaluation d'un test sont valides (validité interne), la question de l'utilisation de ce test dans son cabinet (validité externe) se pose. Il faut alors se poser les questions suivantes (Tableau 1).

*– La reproductibilité et l'interprétation des résultats du test sont-elles données dans mon contexte clinique?*

Une caractéristique importante d'un test diagnostique à connaître est sa reproductibilité. Ce test doit donner les mêmes résultats chez des patients stables ou dans les mêmes conditions si le test est répété plusieurs fois. Si un résultat de test est difficilement reproductible, cela peut tenir à des problèmes techniques ou méthodologiques, ou à la subjectivité de l'interprétation de ce test. C'est pour cela qu'il faut avoir des indications sur la reproductibilité. Ce qui joue un rôle surtout lorsque l'interprétation du test exige des connaissances particulières de l'examineur, comme l'interprétation de tomographies computerisées, d'examens aux ultrasons, etc. Si la reproductibilité d'un test n'est que moyenne, mais s'il permet tout de même de faire la distinction entre résultats pathologiques et normaux, il y a peu de raisons de s'inquiéter. Si la reproductibilité d'un test est très élevée, et si l'appréciation des examinateurs ne varie que très peu, cela signifie soit que ce test est simple à utiliser, soit que ceux qui l'interprètent sont très qualifiés. Dans ce dernier cas, avec des interprètes moins qualifiés, le même test sera peu reproductible dans un autre environnement clinique.

Il n'y a aucune indication sur la reproductibilité dans notre exemple. Nous vous renvoyons à la littérature pour les détails techniques. L'interprétation du test semi-quantitatif est simplifiée en positif et négatif selon des critères bien définis. D'après notre expérience, cela correspond à la réalité de la pratique clinique.

*– Est-il possible d'estimer une probabilité prétest cliniquement raisonnable pour la maladie cible de mon patient?*

L'estimation de la probabilité prétest est un point très important dans la pratique de tout clinicien. Nous pouvons nous baser sur notre expérience, mais notre mémoire de patients semblables est très souvent unilatérale, car nous nous souvenons plutôt des diagnostics man-

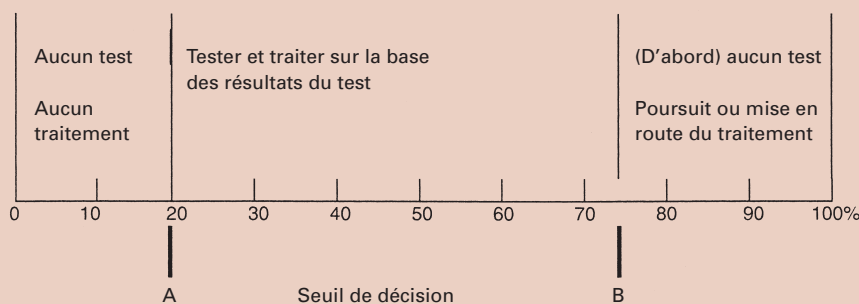
qués, ou sommes guidés par la crainte de passer à côté de rares maladies traitables. Si les caractéristiques du collectif d'une étude le permettent pour nos patients, nous pouvons aussi partir des probabilités prétest de ce collectif.

L'étude de notre exemple présente une autre possibilité. Les auteurs ont utilisé pour leur estimation de la probabilité prétest un score validé de Wells et al. (Tableau 4) [12]. Des critères anamnestiques et cliniques importants pour l'estimation de la probabilité d'une TVP sont rassemblés dans un système de points. Ce score clinique doit naturellement être également considéré comme test avec sensibilité, spécificité et likelihood ratios. Si de nombreux critères sont satisfaits, le «Wells Test» atteint un chiffre élevé. La probabilité prétest de 25,5% (prévalence dans le collectif examiné par Wells) passe à une probabilité posttest de 80%. A l'inverse, avec un nombre de points inférieur, comme dans l'exemple de notre scénario clinique, la probabilité prétest tombe à une probabilité posttest de 3–10%. Et c'est la probabilité prétest pour le prochain test, le dosage des D-dimères.

Une évaluation critique des critères de Wells dépasse le cadre de ce travail, mais les critères de «critical appraisal» d'une règle décisionnelle («decision rule») existent [7, 13]. La règle décisionnelle de Wells, avec la prévalence indiquée dans l'étude, sont des directives utilisables pour estimer la probabilité prétest d'une TVP.

*– La probabilité posttest est-elle dans un domaine qui va modifier la prise en charge de mon patient?*

Si le test diagnostique est négatif, ou que la likelihood ratio est de 0,1, la probabilité posttest peut être si faible que nous pouvons exclure une maladie. Le résultat négatif du test nous a conduit en dessous du seuil du test auquel nous aurions envisagé d'autres examens diagnostiques (Figure 2). A l'inverse, avec une probabilité moyenne sur la base de notre hypothèse diagnostique, nous mettrions en route un traitement, si nécessaire, tout en cherchant à préciser le diagnostic après coup. Avec une forte



**Figure 2.** Probabilité prétest et seuil décisionnel d'après Sackett et al. [7].

probabilité posttest pour la maladie cible, selon la clinique et la maladie, il se peut qu'aucun diagnostic complémentaire ne soit nécessaire et que nous commencions immédiatement un traitement, ou que nous le poursuivions.

### Dénouement du scénario

Selon les critères de Wells (Tableau 4), vous estimez que la probabilité prétest de TVT est faible, de 3 à 10% [12, 14]. Dans la population ambulatoire de l'étude de l'Hôpital cantonal de Bâle, la prévalence de la TVP est de 20,8%. Vous trouvez une probabilité prétest de 3 à 10% sur l'échelle de gauche du nomogramme de Fagan. Sur l'échelle du milieu, vous recherchez la likelihood ratio de 0,23 et lisez sur l'échelle de droite une probabilité posttest de 1 à 2,5%. Avec un dosage des D-dimères négatifs, la probabilité que notre patiente ait une TVP est d'environ

1%, 2,5% dans le pire des cas. Cet exemple illustre qu'à une likelihood ratio donnée, le bénéfice du test dépend de la probabilité prétest. Plus elle est grande, plus le diagnostic sera confirmé à une likelihood ratio donnée. A l'inverse, plus la probabilité prétest est faible, plus il sera possible d'exclure un diagnostic à une likelihood ratio donnée.

Il est maintenant minuit. Vous êtes convaincu de n'avoir pas raté une thrombose chez votre patiente de 27 ans, avec une probabilité prétest faible et un test des D-dimères négatif, et vous allez pouvoir dormir tranquille avec un «risque résiduel» d'environ 2%. Vous avez convoqué votre patiente pour le lendemain pour voir comment elle évolue («test of time»). La sécurité absolue en médecine n'existe en fait que dans les têtes des juristes. Vous avez réservé la prochaine soirée de libre pour aller au concert avec votre partenaire.

**Tableau 4.**  
**Modèle clinique permettant d'estimer les probabilités prétest d'une thrombose veineuse profonde (TVP) selon Wells [12].**

Prévalence de la TVP dans le collectif de Wells: 25,5%

Symptôme clinique	Score
Maladie cancéreuse active (sous traitement, traitement palliatif ou traitement pendant les 6 mois précédents)	1
Paralysie, parésie, ou immobilisation dans une attelle des membres inférieurs	1
Alitement récent pour plus de 3 jours, ou intervention chirurgicale majeure au cours des 4 semaines précédentes	1
Tension localisée le long des veines profondes	1
Toute la jambe enflée	1
Circonférence des mollets >3 cm par rapport à la jambe asymptomatique (mesurée 10 cm en dessous de la tubérosité tibiale)	1
Œdème prenant le godet (plus marqué qu'au niveau de la jambe asymptomatique)	1
Présence de veines collatérales superficielles (pas de varices)	1
Autre diagnostic aussi probable ou plus probable qu'une TVP	-2

Appréciation:

Nombre de points	Probabilité d'une TVP	Probabilité (%)
0 ou <0	Minime	3-10
1 ou 2	Moyenne	17-33
3 ou plus	Forte	75-85

### Références

1. Aus der Au C, Wuillemin WA. D-Dimer-Test bei Verdacht auf venöse Thromboembolie. *Ars Medici* 1999; 11:751.
2. Farrell S, Hayes T, Shaw M. A negative SimpliRED D-dimer assay result does not exclude the diagnosis of deep vein thrombosis or pulmonary embolus in emergency department patients. *Ann Emerg Med* 2000;35:121-5.
3. Anderson DR, Wells PS, Stiell I, MacLeod B, Simms M, Gray L, et al. Management of patients with suspected deep vein thrombosis in the Emergency Department: Combining use of a clinical diagnosis model with D-dimer testing. *J Emerg Med* 2000;19:225-30.
4. Evidence-Based Medicine. BMJ Publishing Group; 2000. Vol. 5; 3; 92-93
5. Aschwanden M, Labs KH, Jeanne- ret C, Gehrig A, Jaeger KA. The Value of rapid D-dimer testing combined with structured clinical evaluation for the diagnosis of deep vein thrombosis. *J Vasc Surg* 1999;30:929-35.
6. Lennox AF, Konstantinos TD, Serunkuma S, Daskalopoulou SE, Nicolaides AN. Combination of a clinical risk assessment score and rapid whole blood D-dimer testing in diagnosis of deep vein thrombosis in symptomatic patients. *J Vasc Surg* 1999;30:794-803.
7. Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Hayes RB. Evidence-based Medicine. How to practice and teach EBM. Second Edition New York, Edinburgh, London, Toronto, Philadelphia, St. Louis, Sidney: Churchill Livingstone; 2000.
8. Thomson DMP, Krupey J, Freedman SO, Gold P. The radioimmunoassay of circulating carcino-embryonic antigen of the human digestive system. *Proc Natl Acad Sci USA*. 1969;64:161-7.
9. The PIOPED Investigators Value of ventilation/perfusion scan in acute pulmonary embolism: results of the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED). *JAMA* 1990;263:2751-9.
10. Fagan TJ. Nomogram for Bayes theorem. *N Engl J Med* 1975;293: 257.
11. Pewsner D, Bleuer J, Bucher HC, Battaglia M, Jüni P, Egger M. Derivation auf der Spur? Das Bayes'sche Theorem und die Diagnostik in der Grundversorgung. *Swiss Medical Forum* 2001;3: (in press)
12. Wells PS, Anderson DR, Bormanis J, Guy F, Mitchel M, Gray L, et al. Value of assessment of pretest probability of deep-vein thrombosis in clinical management. *Lancet* 1997; 350:1795-8.
13. Richardson WS, Wilson M, Guyatt GH, Cook DJ, Nishikawa J for the Evidence-Based Medicine Working Group. Users' Guide to the medical literature: XV. How to use an article about disease probability for differential diagnosis. *JAMA* 1999; 281: 2114-9.
14. Wells PS, Hirsh J, Anderson D, Lensing A, Fortster G, Kearon C. Accuracy of clinical assessment of deep-vein thrombosis. *Lancet* 1995;345: 1326-30.
15. Bennett NM, Paris MC. Cardiovascular Problems. In: Black ER, Bordley DR; Tape TG; Panzer RJ (editors). Diagnostic strategies for common medical problems. Second Edition Philadelphia, Pennsylvania: American College of Physicians; 1999. pp. 47-60.

### Autre littérature sur ce sujet

- Bucher HC, Schmidt JG, Steurer J. Kritische Beurteilung einer Arbeit zu einem diagnostischen Test. *Schweiz Rundsch Med Prax* 1998;87:1096-102.
- Greenhalg T. How to read a paper. The basics of evidence based medicine. London: MJ Publishing Group; 1997.