



**HAL**  
open science

## The expert panel for Script Concordance Tests: A truly adequate reference?

Luc Dauchet, Raphael Bentegeac, Haress Ghauss, Marc Hazzan, Patrick Truffert, Philippe Amouyel, Victoria Gauthier, Aghiles Hamroun

### ► To cite this version:

Luc Dauchet, Raphael Bentegeac, Haress Ghauss, Marc Hazzan, Patrick Truffert, et al.. The expert panel for Script Concordance Tests: A truly adequate reference?. *La Revue de Médecine Interne*, 2024, *La Revue de Médecine Interne*, Online ahead of print. 10.1016/j.revmed.2024.05.023 . hal-04684848

**HAL Id: hal-04684848**

**<https://hal.univ-lille.fr/hal-04684848v1>**

Submitted on 3 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Disponible en ligne sur

ScienceDirect  
www.sciencedirect.com

Elsevier Masson France

EM|consulte  
www.em-consulte.com



Idées et débats

## Le panel d'experts pour les tests de concordance de script : une référence réellement adéquate ?

### *The expert panel for Script Concordance Tests: A truly adequate reference?*

Luc Dauchet<sup>a,b</sup>, Raphaël Bentegeac<sup>a,b</sup>, Haress Ghauss<sup>a,b</sup>, Marc Hazzan<sup>c</sup>, Patrick Truffert<sup>d</sup>,  
Philippe Amouyel<sup>a,b</sup>, Victoria Gauthier<sup>a,b</sup>, Aghilès Hamroun<sup>a,b,\*</sup>

<sup>a</sup> Service de santé publique, épidémiologie, économie de la santé et prévention, CHU de Lille, 59000 Lille, France

<sup>b</sup> UMR1167 RID-AGE, Institut Pasteur de Lille, Inserm, université de Lille, CHU de Lille, 59000 Lille, France

<sup>c</sup> Service de néphrologie, dialyse, transplantation rénale et aphérèse, hôpital Claude-Huriez, université de Lille, CHU de Lille, 59000 Lille, France

<sup>d</sup> Service de néonatalogie, hôpital Jeanne-de-Flandres, université de Lille, CHU de Lille, 59000 Lille, France

#### ARTICLE INFO

##### Article history:

Available online xxx

##### Mots clés:

Tests de concordance de script

TCS

Raisonnement clinique

Raisonnement bayésien

Raisonnement probabiliste

Panel d'experts

##### Keywords:

Script concordance test

Clinical reasoning

Bayesian reasoning

Probabilistic reasoning

Expert panel

#### RÉSUMÉ

Les tests de concordance de script (TCS) constituent une modalité docimologique introduite par arrêté dans les épreuves dématérialisées en 2024. Leur objectif serait d'évaluer le raisonnement clinique en situation d'incertitude. En pratique, les TCS interrogent l'étudiant sur l'impact d'une nouvelle information sur la probabilité d'une hypothèse formulée a priori à partir d'une situation clinique authentique. Cette démarche s'apparente au raisonnement probabiliste (ou bayésien). En raison de l'incertitude liée à la situation clinique explorée, les TCS ne confrontent pas la réponse de l'étudiant à celle attendue dans un référentiel de connaissances théoriques. C'est la distribution des réponses d'un panel de médecins expérimentés qui sert à établir le barème de la question. Les données de la littérature suggèrent que les médecins, même expérimentés, comme la plupart des humains ont un raisonnement probabiliste intuitif le plus souvent biaisé. Ces biais posent la question de la pertinence d'utiliser les réponses du panel d'experts comme barème de notation pour les TCS.

© 2024 The Author(s). Published by Elsevier Masson SAS on behalf of Société Nationale Française de Médecine Interne (SNFMI). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

#### ABSTRACT

The Script Concordance Tests (SCTs) are an examination modality introduced by decree in the French National Ranking Exam for medical students in 2024. Their objective is to evaluate clinical reasoning in situations of uncertainty. In practice, SCTs assess the impact of new information on the probability of a hypothesis formulated a priori based on an authentic clinical scenario. This approach resembles probabilistic (or Bayesian) reasoning. Due to the uncertainty associated with the explored clinical situation, SCTs do not compare the student's response to an expected one in a theoretical knowledge reference. Instead, the distribution of responses from a panel of experienced physicians is used to establish the question's scoring scale. Literature data suggest that physicians, even experienced ones, like most humans, often exhibit biased intuitive probabilistic reasoning. These biases raise questions about the relevance of using expert panel responses as scoring scales for SCTs.

© 2024 The Author(s). Published by Elsevier Masson SAS on behalf of Société Nationale Française de Médecine Interne (SNFMI). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\* Auteur correspondant.

E-mail address: [aghiles.hamroun@univ-lille.fr](mailto:aghiles.hamroun@univ-lille.fr) (A. Hamroun).

<https://doi.org/10.1016/j.revmed.2024.05.023>

0248-8663/© 2024 L'Auteur(s). Publié par Elsevier Masson SAS au nom de Société Nationale Française de Médecine Interne (SNFMI). Cet article est publié en Open Access sous licence CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Please cite this article as: L. Dauchet, R. Bentegeac, H. Ghauss et al., Le panel d'experts pour les tests de concordance de script : une référence réellement adéquate ? Rev Med Interne, <https://doi.org/10.1016/j.revmed.2024.05.023>

## 1. Introduction

En France, la réforme du second cycle et des modalités d'accès au troisième cycle des études médicales vise à développer l'apprentissage du raisonnement clinique, et non pas les seules connaissances médicales, pour mieux former les étudiants à leurs futures fonctions de professionnel de santé [1]. Le raisonnement clinique est très complexe à conceptualiser. Sa théorisation renvoie à l'intervention des connaissances théoriques associée à une multitude de processus cognitifs, que l'on regroupe le plus souvent au sein de deux catégories, dites « processus analytiques » et « processus intuitifs » [2]. Les processus analytiques correspondent à une succession d'étapes faisant intervenir un cheminement conscient actif au cours duquel le professionnel récolte de façon systématique un certain nombre d'informations sur lesquelles s'appuyer pour générer des hypothèses. À l'inverse, les processus intuitifs seraient réalisés de façon passive, en mobilisant de manière tacite des hypothèses correspondant à la reconnaissance d'exemples concrets ou d'expériences antérieures similaires au cas rencontré (également appelées des scripts) [3]. L'ensemble de ces théories se rejoignent sur le rôle central joué par la reconnaissance intuitive des scripts, précisément cible de l'évaluation par les tests de concordance de script (TCS) [4].

## 2. Les TCS, un pont naturel entre raisonnement clinique et raisonnement probabiliste ?

En pratique, les TCS permettraient à l'étudiant d'explicitier certaines étapes clés de son raisonnement clinique [5]. Ils évalueraient ainsi la capacité de l'étudiant à prendre des décisions pertinentes en situation d'incertitude, en comparant ses réponses à celles d'un panel d'experts. Les TCS se composent de deux éléments : une vignette clinique courte et des questions (généralement au nombre de 3). Chaque question offre à l'étudiant la possibilité d'exprimer sur une échelle de Likert à 5 niveaux l'impact d'une nouvelle information sur la probabilité de l'hypothèse formulée a priori à partir de la vignette clinique [6] (Fig. 1).

Lors de l'expérience de rédaction et de réponse aux TCS en tant que panélistes, nous avons constaté un lien évident entre la structure des TCS et le « raisonnement probabiliste ». En effet, le raisonnement probabiliste, également appelé raisonnement bayésien, est un type de raisonnement inductif qui quantifie l'incertitude par les probabilités, en utilisant la loi des probabilités conditionnelles de Bayes [7]. De manière plus pratique, le raisonnement probabiliste peut être considéré comme l'action de changer d'avis, avec une confiance dans une hypothèse a priori et la mise à jour de cette confiance après avoir observé une nouvelle information. En situation clinique, l'hypothèse formulée est très souvent celle d'une maladie, et l'information peut être n'importe quel élément de sémiologie : un signe clinique, un symptôme, le résultat d'un test diagnostique, un antécédent, ou autre. Une nouvelle information pertinente peut ainsi modifier notre confiance dans cette hypothèse, qui va ainsi évoluer en hypothèse a posteriori. Si l'on

s'intéresse par exemple à une patiente de 72 ans adressée en consultation pour altération de l'état général, douleurs articulaires diffuses et syndrome inflammatoire biologique d'étiologie inexplicquée, le clinicien peut imaginer plusieurs hypothèses à partir du simple courrier d'adressage du médecin traitant. Il pourrait par exemple évoquer une maladie de Horton avec une probabilité relativement faible compte tenu de la faible prévalence de la maladie, y compris chez les patients de plus de 50 ans : on parlera ici de la probabilité a priori. En revanche, si le médecin met en évidence à l'anamnèse des douleurs au niveau des mâchoires à la mastication ou une hypersensibilité du cuir chevelu lors du coiffage, la probabilité d'une maladie de Horton serait ainsi revue à la hausse : on parlera ici de la probabilité diagnostique a posteriori.

Bien que le raisonnement bayésien soit pertinent pour illustrer la démarche de raisonnement clinique, il reste néanmoins peu appliqué dans la pratique clinique [8-10]. En effet, les cliniciens ont peu recours au raisonnement probabiliste dans leur pratique de tous les jours, entre 3 et 8 % des médecins interrogés selon les études [8,10]. En plus d'être décrite comme une méthode peu intuitive, d'autres raisons ont été évoquées pour expliquer cette réticence à recourir au raisonnement probabiliste, telles que le caractère très consommateur en ressources cognitives, le manque de formation, ainsi que différents biais cognitifs (biais de négligence de la probabilité a priori, d'ancrage et de confirmation notamment) [9,10]. À l'inverse, bien que les indicateurs de performance diagnostique (sensibilité, spécificité et valeurs prédictives) soient plus souvent utilisés par les cliniciens (environ 50 % des répondants dans les enquêtes), leur application et interprétation sont très souvent erronées [8,11].

Ainsi, lorsqu'un TCS questionne l'apport d'une nouvelle information pour renforcer ou non une hypothèse, il interroge par définition une démarche correspondant au raisonnement bayésien. Bien que les TCS semblent de fait explorer le raisonnement bayésien, aucun des référentiels cités par les recommandations HAS [12] ne fait mention du raisonnement probabiliste ou du théorème de Bayes [5,6,13-16]. *In fine*, le score de l'étudiant est déterminé en confrontant sa réponse à la distribution des modalités de réponse d'un panel d'experts. C'est ainsi le degré de concordance entre la réponse de l'étudiant et celle du panel d'experts qui est évalué (Fig. 2). Compte tenu du rôle central joué par le panel d'experts dans la construction du barème des TCS, il est nécessaire de questionner sa pertinence.

## 3. Le panel d'expert en tant que référence : un excès de confiance au regard de ses performances ?

C'est la modalité de réponse la plus fréquemment retenue par les panélistes qui permettra à l'étudiant d'obtenir le maximum de points (Fig. 2). Le TCS repose sur le postulat que la réponse modale des experts est la plus pertinente. Ce présupposé n'a à notre connaissance jamais fait l'objet d'une réelle évaluation. Cette évaluation serait très difficile à réaliser compte tenu de l'absence, par définition, d'un référentiel définissant la « bonne réponse » à un

Une patiente âgée de 65 ans est adressée aux urgences pour suspicion d'embolie pulmonaire. Elle a des antécédents de diabète, d'insuffisance cardiaque et de BPCO.						
Si vous comptiez prescrire :	Et qu'alors vous apprenez :	Cette action deviendrait :				
(1) Un angioscanner thoracique	Que la patiente est sous metformine	-2	-1	0	+1	+2
(2) Un dosage des D-dimères	Que la patiente est actuellement traitée pour un cancer ovarien	-2	-1	0	+1	+2
(3) Un échodoppler veineux des membres inférieurs	Que la patiente présente un signe de Homans positif	-2	-1	0	+1	+2

Fig. 1. Exemple de test de concordance de script (adapté de Fournier JP et al. [6]).

Une patiente âgée de 65 ans est adressée aux urgences pour suspicion d'embolie pulmonaire. Elle a des antécédents de diabète, d'insuffisance cardiaque et de BPCO.						
Si vous comptiez prescrire :	Et qu'alors vous apprenez :	Cette action deviendrait :				
(1) Un angioscanner thoracique	Que la patiente est sous metformine	-2	-1	0	+1	+2
<i>Réponse des experts</i>		0	0	3	4*	2
(2) Un dosage des D-dimères	Que la patiente est actuellement traitée pour un cancer ovarien	-2	-1	0	+1	+2
<i>Réponse des experts</i>		5*	3	1	0	0
(3) Un échodoppler veineux des membres inférieurs	Que la patiente présente un signe de Homans positif	-2	-1	0	+1	+2
<i>Réponse des experts</i>		0	0	5*	3	1

-2 : complètement inutile  
 -1 : inutile  
 0 : ni plus ni moins utile  
 +1 : plus utile  
 +2 : très utile

Calcul du barème à partir des réponses formulées par le panel d'experts (n = 9)					
	-2	-1	0	+1	+2
<b>Hypothèse (1)</b>	0/4 = 0 point	0/4 = 0 point	3/4 = 0,75 point	4/4 = 1 point	2/4 = 0,5 point
<b>Hypothèse (2)</b>	5/5 = 1 point	3/5 = 0,6 point	1/5 = 0,2 point	0/5 = 0 point	0/5 = 0 point
<b>Hypothèse (3)</b>	0/5 = 0 point	0/5 = 0 point	5/5 = 1 point	3/5 = 0,6 point	1/5 = 0,2 point

Fig. 2. Détail du calcul du barème à partir de l'exemple de Test de concordance de script (adapté de Fournier JP et al. [6]). \*Le calcul du barème repose sur la réponse modale des experts : cette dernière équivaut à 1 point tandis que la cotation des autres réponses est définie par le nombre de coches par les experts rapporté au nombre de coches de la réponse modale.

TCS. Malgré l'absence d'évaluation formelle sur le sujet, on peut raisonnablement considérer que les experts ont les compétences et connaissances requises pour un tel objectif. Cependant, certains éléments peuvent remettre en cause la validité de ce présupposé [9,11,17,18]. Rappelons à ce titre que l'on interroge les experts, de la même manière que les étudiants, sur un exercice très particulier et complexe d'estimation de probabilités conditionnelles. Ce type d'exercice renvoie notamment au raisonnement probabiliste dont les fondements théoriques sont assez méconnus de la communauté médicale [11].

#### 4. Des experts peu coutumiers du raisonnement probabiliste

Les données de la littérature suggèrent que les humains en général sont sujets à de nombreux types de biais à l'occasion d'un raisonnement probabiliste ou bayésien. Une étude menée en 2020 par une équipe de psychologues de l'Université de Harvard a montré que l'inférence probabiliste humaine déviait systématiquement du raisonnement optimal [19]. Dans une stratégie d'optimisation de ressources cognitives, l'humain évite de réaliser un raisonnement probabiliste pour se concentrer sur une reconnaissance de la distribution des probabilités d'un énoncé quel qu'il soit. En pratique, à partir d'un énoncé portant sur l'estimation de la probabilité d'être malade avec un test diagnostique positif et le détail de ses performances diagnostiques, le cerveau humain aura spontanément tendance à estimer d'emblée la probabilité finale d'être malade plutôt que de décomposer sa réflexion en trois étapes constitutives du raisonnement probabiliste : (1) quelle est la probabilité pré-test (équivalente à la prévalence de la maladie en l'absence d'autres informations) ? (2) quelle est la valeur diagnostique de ce test positif (équivalente à son rapport de vraisemblance positif) ? (3) quelle est la probabilité post-test dans cette situation (tenant compte à la fois de la probabilité pré-test et du rapport de vraisemblance) ? Ce raisonnement, modélisé mathématiquement par le théorème de Bayes, peut être appliqué en clinique à l'aide du nomogramme de Fagan [7], dont l'utilisation et l'interprétation

sont au programme du deuxième cycle des études médicales et des épreuves dématérialisées nationales (Item 3. Le raisonnement et la décision en médecine, rang A). Si l'on prend l'exemple d'un patient qui tousse sans fièvre ni notion de contagion en pleine période épidémique, sa probabilité d'être atteint de la COVID-19 est a priori faible  $p(M) = 10\%$ . On apprend ensuite que le test antigénique, dont le rapport de vraisemblance positif est très élevé ( $RV_+ = 90$ ), est revenu positif. On peut estimer une probabilité a posteriori pour ce patient d'être atteint du COVID. Pour réaliser cette estimation à l'aide du nomogramme, il suffit de tracer une droite partant de la probabilité initiale sur l'échelle verticale de gauche et passant par le rapport de vraisemblance rapporté sur l'échelle verticale centrale (Fig. 3). Le point d'intersection avec l'échelle verticale de droite correspond ainsi à la probabilité a posteriori, ici 91%. En clinique, les difficultés des médecins dans l'application du raisonnement probabiliste ont été largement décrites dans la littérature : dans une revue systématique publiée en 2015, incluant 24 études portant sur la compréhension et l'interprétation des indicateurs de performances diagnostiques d'un test par des cliniciens, les auteurs concluaient à une mauvaise compréhension de ces paramètres induisant très fréquemment une estimation erronée de la probabilité post-test, le plus souvent très largement à la hausse [17]. Fait important à noter, le mode de présentation des énoncés, sous forme de probabilités conditionnelles, en fournissant la valeur des indices de performance diagnostique (sensibilité, spécificité) ou en faisant intervenir des énoncés intuitifs, voire une représentation visuelle (échelle visuelle analogique par exemple), a un impact direct sur la performance des cliniciens dans l'estimation de la probabilité a posteriori [17]. Dans une étude réalisée en Espagne en 2013, Garcia-Retamero R. et Hoffrage U. ont proposé à 81 médecins et 81 patients de réaliser des inférences diagnostiques à partir d'énoncés simples incluant des informations sur la prévalence, la sensibilité et le taux de faux-positifs d'un test diagnostique pour une maladie définie telle que la mammographie pour le cancer du sein [20]. Le même énoncé était proposé en intégrant tantôt des probabilités (« la prévalence de la maladie au sein de la population est de 0,4% »), tantôt des fréquences naturelles (« au sein de la population,

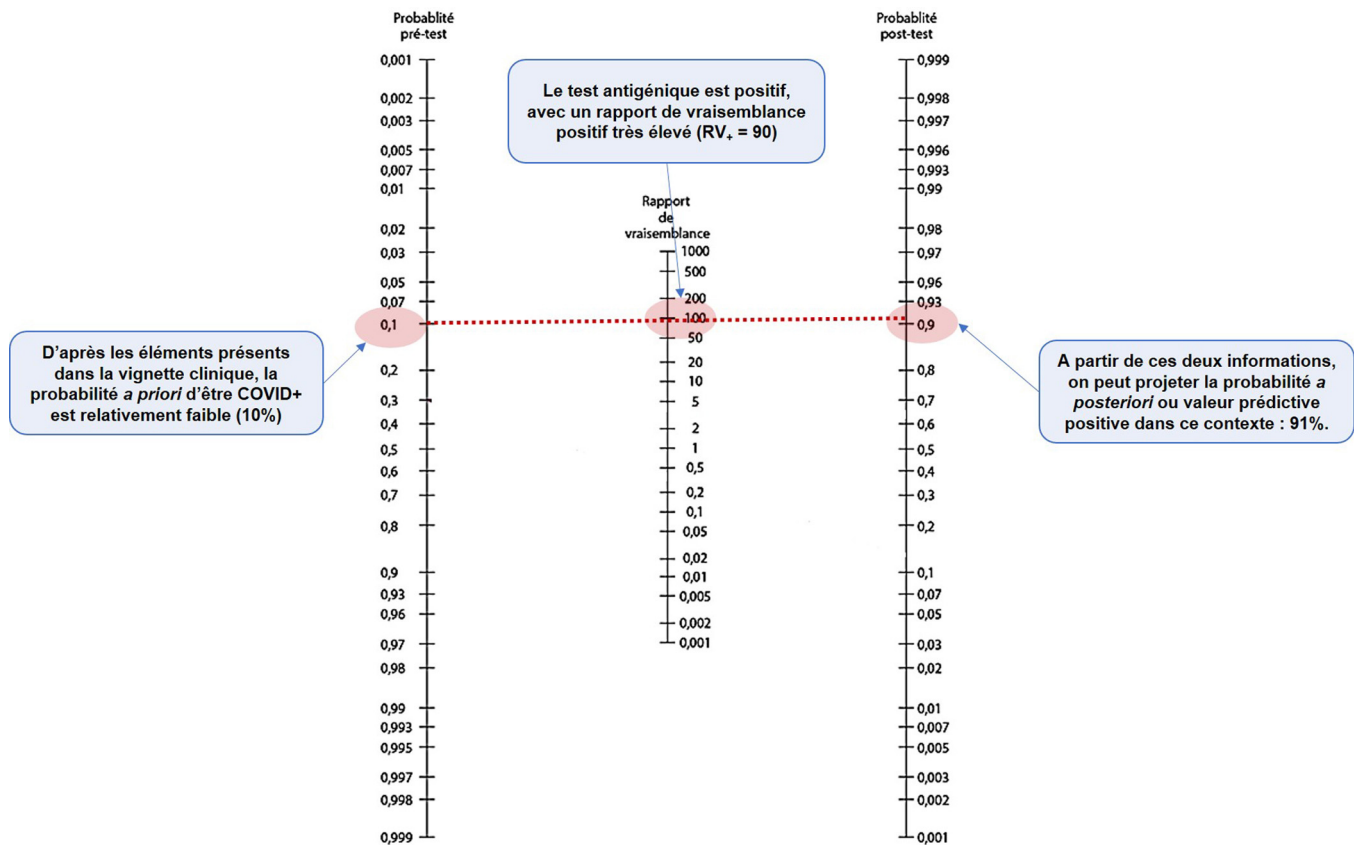


Fig. 3. Application pratique d'un raisonnement probabiliste à partir d'un exemple clinique (test antigénique COVID) : illustration par le nomogramme de Fagan.

40 personnes sur 10 000 sont malades »). La moitié de la population disposait également d'une aide visuelle infographique représentant l'information numérique. Les inférences diagnostiques apparaissaient plus souvent justes lorsque l'énoncé intégrait des fréquences naturelles plutôt que des probabilités (48 % contre 22 % chez les médecins, 25 % contre 5 % chez les patients) et les performances étaient boostées par l'apport d'une aide visuelle (près de 70 % de réponses correctes chez les médecins, et plus de 50 % chez les patients).

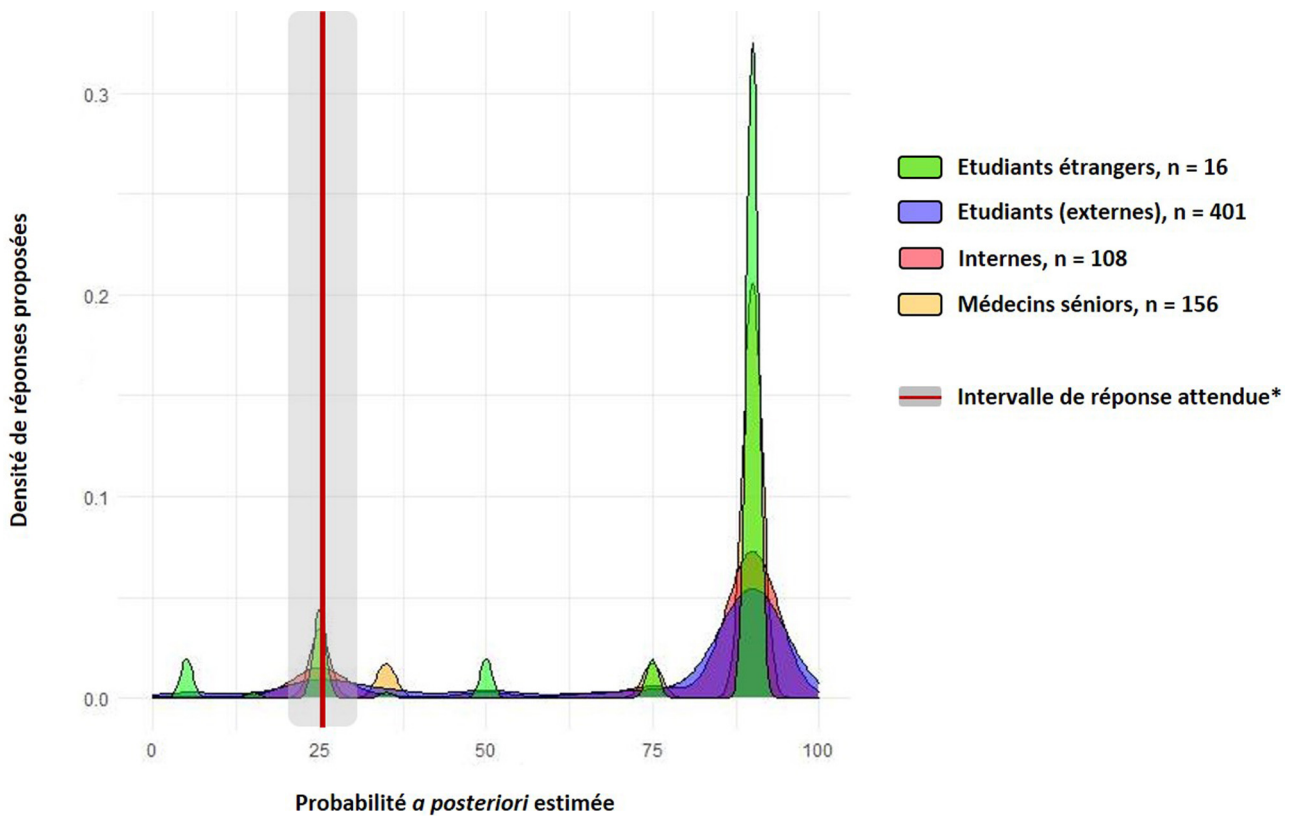
Au regard de la littérature, on peut s'interroger sur la pertinence du panel d'experts pour établir le barème de notation des TCS, dans la mesure où le raisonnement probabiliste est sujet à différents biais résultant le plus souvent dans une surestimation de la probabilité a posteriori. Un étudiant qui serait performant en matière de raisonnement probabiliste pourrait être désavantagé car s'éloignant de la réponse modale des panélistes, en dépit d'une réponse plus pertinente. Une étude réalisée en France a évalué l'estimation par 681 étudiants, internes et médecins séniors, de probabilités conditionnelles à partir des indices de performance de tests diagnostiques, exercice en théorie accessible à partir des connaissances du deuxième cycle des études médicales [11]. La question posée aux participants consistait à estimer la probabilité qu'un patient soit atteint de la COVID-19 sachant qu'il est positif au test diagnostique, à partir des informations suivantes : la prévalence de la maladie est de 0,4 %, la sensibilité du test est de 90 % et sa spécificité de 99 %. Au total, 85 % des médecins interrogés ont échoué à estimer la probabilité post-test à partir de la prévalence et du rapport de vraisemblance du test, avec un haut degré de confiance dans leur réponse (en médiane 80 sur 100) (Fig. 4). La courbe de densité représentée en Fig. 4 assimilable à un « histogramme avec une infinité de bâtons » met en évidence un pic pour les probabilités situées entre 80 et 100 %, concentrant

ainsi la très grande majorité des réponses des participants (environ 70 %), qu'ils soient étudiants ou médecins séniors. In fine, seuls 11 % des répondants ont indiqué une probabilité estimée entre 20 et 30 % pour ce patient d'être atteint de la COVID-19, correspondant à une fourchette acceptable autour de la réponse attendue à 26 %. Cette étude confirme les faibles performances des médecins dans l'utilisation des indices informationnels des tests diagnostiques tels que la sensibilité ou la spécificité pour l'estimation d'une probabilité a posteriori avec toujours la même tendance à la surestimation dans l'inférence diagnostique.

### 5. Une méconnaissance du raisonnement probabiliste potentiellement compensée par l'expérience ?

Dans une étude publiée en 2021, rapportant une expérimentation aux contours proches du format des TCS, 723 médecins exerçant dans huit états des États-Unis ont été confrontés à 4 scénarios fréquemment rencontrés en soins primaires [9]. Pour chacun d'entre eux, une hypothèse diagnostique était évoquée, à savoir une pneumopathie aiguë communautaire, une ischémie myocardique, un cancer du sein, ou une infection urinaire. L'objectif était d'évaluer la capacité de ces médecins séniors à estimer la probabilité pré-test sur la base des éléments de contexte présents dans la vignette clinique, puis la probabilité post-test ainsi que le rapport de vraisemblance du test proposé (radiographie thoracique, épreuve d'effort, mammographie et examen cytotabac-tériologique des urines, respectivement). Les scénarios choisis avaient l'avantage de pouvoir se confronter à des probabilités tirées des données objectives de la littérature. La conclusion de cette étude est que les probabilités post-test ont été systématiquement largement surévaluées par les cliniciens quels que soient le scénario et le résultat du test (négatif ou positif). Si l'on prend





**Fig. 4.** Répartition de la densité des réponses proposées selon le niveau d'expérience au sein du panel interrogé (résultats extraits de Lakhli et al. [11]). \*Toutes les réponses situées dans l'intervalle  $26 \pm 5\%$  ont été considérées comme correctes.

l'exemple de la probabilité de cancer du sein après une mammographie positive chez une femme de 45 ans asymptomatique et sans facteur de risque : les internes et les médecins séniors estimaient la probabilité de cancer du sein en moyenne entre 50 et 60 % pour une probabilité attendue de 3–9 % d'après les données de la littérature. Cette surestimation de la probabilité post-test était en grande partie liée à une surestimation de la probabilité pré-test et de la valeur estimée des rapports de vraisemblance pour l'ensemble des scénarios proposés. De façon surprenante, le niveau d'expérience des médecins participant n'était pas associé à la validité des estimations, avec pour certains scénarios une estimation davantage biaisée chez les médecins séniors que chez les internes moins expérimentés. De la même manière, dans l'étude réalisée par Lakhli C. et al. en France, la concordance entre le degré de confiance annoncé et l'exactitude des réponses était associée au niveau d'expertise du répondant : alors que les étudiants les moins performants présentent une concordance élevée (les erreurs sont associées à un faible degré de confiance et inversement les réponses correctes sont associées à un degré de confiance plus élevé), celle-ci baisse à mesure que le répondant est expert (aucune relation entre la justesse de la réponse et le degré de confiance annoncé) [11].

## 6. Conclusion

Les TCS, dont l'objectif est d'explorer le raisonnement clinique en situation d'incertitude, mettent à l'épreuve les principes du raisonnement probabiliste. Dans leur conception, c'est le panel d'experts constitué de médecins séniors qui fait office de référence afin d'établir le barème de notation. Alors que le raisonnement des cliniciens intègre en pratique une multitude d'informations, les TCS les placent dans une situation clinique théorique inhabituelle, circonscrite à quelques informations à partir desquelles ils doivent estimer précisément une probabilité pré-test, un rapport

de vraisemblance, ainsi qu'une probabilité post-test. En pratique clinique, l'inférence diagnostique du clinicien tient compte de l'ensemble des informations disponibles à l'heure de porter le diagnostic alors que les TCS les contraignent à une décomposition artificielle du raisonnement bayésien à l'orée de chaque nouvelle information prise de façon individuelle. Les données de la littérature questionnent ainsi la pertinence du panel d'experts pour servir de référence, et peut-être plus globalement la pertinence même des TCS dans l'évaluation des étudiants en médecine.

## Déclaration de liens d'intérêts

Les auteurs déclarent ne pas avoir de liens d'intérêts.

## Contribution des auteurs

Luc Dauchet, Raphaël Bentegeac, Haress Ghauss, Aghiles Hamroun : conceptualisation, bibliographie, réalisation des figures, écriture de la première version du manuscrit ; Victoria Gauthier, Marc Hazzan, Patrick Truffert, Philippe Amouyel : relecture et édition du manuscrit ; Marc Hazzan, Patrick Truffert, Philippe Amouyel : supervision ; l'ensemble des auteurs a validé le contenu du manuscrit dans sa version finale.

## References

- [1] Cohen Aubart F, Roux D. La réforme du deuxième cycle des études médicales en France : risque ou opportunité ? *Rev Med Interne* 2021;42(3):149–53.
- [2] Charlin B, Lubarsky S, Millette B, et al. Clinical reasoning processes: unravelling complexity through graphical representation. *Med Educ* 2012;46(5):454–63.
- [3] Pelaccia T, Tardif J, Tribi E, Charlin B. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Med Educ Online* 2011;16, <http://dx.doi.org/10.3402/meo.v16i0.5890> [PMID: 21430797; PMCID: PMC3060310].

- [4] Charlin B, Tardif J, Boshuizen HP. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med* 2000;75(2):182–90.
- [5] Dory V, Gagnon R, Vanpee D, Charlin B. How to construct and implement script concordance tests: insights from a systematic review. *Med Educ* 2012;46(6):552–63.
- [6] Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. *BMC Med Inform Decis Mak* 2008;8:18.
- [7] Hornberger J. Introduction to Bayesian reasoning. *Int J Technol Assess Health Care* 2001;17(1):9–16.
- [8] Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998;104(4):374–80.
- [9] Morgan DJ, Pineles L, Owczarzak J, et al. Accuracy of practitioner estimates of probability of diagnosis before and after testing. *JAMA Intern Med* 2021;181(6):747–55.
- [10] Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. *BMJ* 2009;338:b946, <http://dx.doi.org/10.1136/bmj.b946>.
- [11] Lakhilfi C, Lejeune FX, Rouault M, Khamassi M, Rohaut B. Illusion of knowledge in statistics among clinicians: evaluating the alignment between objective accuracy and subjective confidence, an online survey. *Cogn Res Princ Implic* 2023;8(1):23.
- [12] Haute Autorité de santé (HAS). Test de concordance de script (TCS). Published 2017. Accessed March 19, 2024. [https://www.has-sante.fr/upload/docs/application/pdf/2017-11/test.de.concordance.de\\_script.pdf](https://www.has-sante.fr/upload/docs/application/pdf/2017-11/test.de.concordance.de_script.pdf).
- [13] Sibert L, Fournier JP, Charlin B. In: ECNi épreuve TCS Test de concordance de script. Le guide méthodologique. France: Maloine; 2015.
- [14] Caire F, Sol JC, Charlin B, Isodiri P, Moreau JJ. Le test de concordance de script (TCS) comme outild'évaluation formative des internes en neurochirurgie: implantation du test sur Internet à l'échelle nationale. *Pedag Med* 2004;5(2):87–94.
- [15] Giet D, Massart V, Gagnon R, Charlin B. Le test de concordance de script en 20 questions. *Pedag Med* 2013;14(1):39–48.
- [16] Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE guide no 75. *Med Teach* 2013;35(3):184–93.
- [17] Whiting PF, Davenport C, Jameson C, et al. How well do health professionals interpret diagnostic information? A systematic review. *BMJ Open* 2015;5(7):e008155.
- [18] Ahmadi SF, Khoshkish S, Soltani-Arabshahi K, et al. Challenging script concordance test reference standard by evidence: do judgments by emergency medicine consultants agree with likelihood ratios? *Int J Emerg Med* 2014;7(1):34.
- [19] Dasgupta I, Schulz E, Tenenbaum JB, Gershman SJ. A theory of learning to infer. *Psychol Rev* 2020;127(3):412–41.
- [20] Garcia-Retamero R, Hoffrage U. Visual representation of statistical information improves diagnostic inferences in doctors and their patients. *Soc Sci Med* 2013;83:27–33.