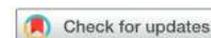


Selecting a risk assessment tool to use in practice: a 10-point guide



Seena Fazel,¹ Achim Wolf²

¹Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK; ²Research Department, St. Andrew's Healthcare, Northampton, UK

Correspondence to Professor Seena Fazel, Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford OX3 7JX, UK; seena.fazel@psych.ox.ac.uk

ABSTRACT

With the increase in the number of risk assessment tools and clinical algorithms in many areas of science and medicine, this Perspective article provides an overview of research findings that can assist in informing the choice of an instrument for practical use. We take the example of violence risk assessment tools in criminal justice and forensic psychiatry, where there are more than 200 such instruments and their use is typically mandated. We outline 10 key questions that researchers, clinicians and other professionals should ask when deciding what tool to use, which are also relevant for public policy and commissioners of services. These questions are based on two elements: research underpinning the external validation, and derivation or development of a particular instrument. We also recommend some guidelines for reporting drawn from consensus guidelines for research in prognostic models.

Fazel S, Wolf A.

Choisir un outil d'évaluation des risques à utiliser dans la pratique : un guide en 10 points

Evid Based Mental Health (2018) – <https://ebmh.bmj.com/content/ebmental/21/2/41.full.pdf>

Résumé

Avec l'augmentation du nombre d'outils d'évaluation des risques et d'algorithmes cliniques dans de nombreux domaines de la science et de la médecine, cet article de perspective présente une vue d'ensemble des résultats de recherche qui peuvent contribuer à éclairer le choix d'un instrument pour une utilisation pratique. Nous prenons l'exemple des outils d'évaluation du risque de violence dans la justice pénale et la psychiatrie légale, où il existe plus de 200 instruments de ce type et où leur utilisation est généralement obligatoire. Nous présentons 10 questions clés que les chercheurs, les cliniciens et les autres professionnels devraient se poser lorsqu'ils décident de l'outil à utiliser, et qui sont également pertinentes pour les politiques publiques et les responsables des services. Ces questions s'appuient sur deux éléments : la recherche qui sous-tend la validation externe, et la dérivation ou le développement d'un instrument particulier. Nous recommandons également quelques lignes directrices pour la rédaction de rapports, tirées des lignes directrices de consensus pour la recherche sur les modèles de pronostic.

La justice pénale a progressivement commencé à évaluer les traitements et les interventions de manière plus rigoureuse, notamment en menant des essais contrôlés randomisés (ECR). Cependant, un domaine a pris du retard, l'évaluation des risques de violence et, en particulier, de récidive¹. La plupart des pays à revenu élevé disposent d'un ensemble d'outils pour évaluer le risque de récidive au moment du prononcé de la peine, des audiences de libération conditionnelle, de la probation ou de la sortie de prison pour déterminer la prise en charge postpénale.

¹ douglas t, Pugh J, Singh I, et al. Risk assessment tools in criminal justice and forensic psychiatry: the need for better data. Eur Psychiatry 2017;42:134–7.

Il existe plus de 200 outils de ce type et de nouveaux apparaissent chaque mois². Un exemple récent est celui d'un outil prédictif utilisé dans la justice pénale de l'État de Géorgie aux États-Unis, qui repose sur cinq éléments statiques et peut être complété en ligne³. Sur la base de ces éléments, il fournit une cotation risque faible, moyen ou élevé, sans information sur la signification de ces catégories ou sur la précision de l'outil. En outre, il existe des outils basés sur les besoins qui prennent beaucoup plus de temps, et, à notre connaissance, n'ont pas fait l'objet d'essais contrôlés randomisés pour tester leurs effets sur les résultats. Dans l'ensemble, on ne sait pas si la plupart de ces instructions sur les risques ont été élaborées et validées selon des méthodes fondées sur des données probantes, en partie parce que les organismes de justice pénale et ceux qui travaillent avec eux ne publient pas leurs résultats dans des revues à comité de lecture, et que les informations fournies dans les rapports en ligne se limitent principalement au travail de développement.

Pour prendre deux exemples au Royaume-Uni, l'approche actuelle intitulée DASH pour évaluer les risques de violence domestique est basée sur un modèle apparemment développé par une seule personne sans validation externe⁴. Dans le domaine de la justice pénale, en Angleterre et au Pays de Galles, un outil d'évaluation des risques en 100 points est appliqué à tous les individus condamnés à une peine de plus de 12 mois, et a été adopté dans de nombreux pays. Bien qu'il ait été validé en Angleterre et au Pays de Galles par la même équipe qui a développé l'outil⁵, ce qui constitue une avancée par rapport à de nombreux autres instruments, nous n'avons pas été en mesure d'identifier une validation indépendante et aucune dans des échantillons en dehors de l'Angleterre et du Pays de Galles.

Compte tenu du nombre croissant de ces instruments et de l'évidente nécessité de stratifier les individus qui passent par le système de justice pénale sur la base du risque, nous aimerions proposer une approche permettant d'évaluer la qualité et la base de données sous-jacente d'une évaluation de risque particulière.

Cette approche devrait permettre de revoir les approches existantes et d'envisager de nouveaux instruments. Pour ce faire, nous nous sommes inspirés de guides méthodologiques pour la recherche sur les pronostics⁶, sur les revues systématiques de l'évaluation des risques dans la justice pénale et sur les études primaires sur la performance de ces outils dans des contextes réels^{7,8,9} (analogues aux études d'efficacité clinique lorsqu'il s'agit de déterminer si les ECR s'appliquent à des protocoles de recherche). Sur la base de ces éléments, nous suggérons de poser 10 questions et de les séparer en deux catégories : celles relatives à la validation externe et aux échantillons de développement

² Singh JP, Desmarais SL, Otto RK, et al. The international risk survey: use and perceived utility of structured violence risk assessment tools in 44 countries. In: Singh JP, Bjorkly S, Fazel S, International perspectives on violence risk assessment. New York: OUP, 2016

³ state of Georgia - board of Pardons and Paroles. Parole decision guidelines risk instrument. 2017 <https://webapplications.pap.state.ga.us/guidelines-calculator/grs/home> (accessed 7 Dec 2017)

⁴ richards I. Domestic abuse, stalking and honour based violence calculator. 2017 <http://www.dashriskchecklist.co.uk> (accessed 7 Dec 2017).

⁵ Howard Pd, Dixon L. The construction and validation of the oasis violence predictor: advancing violence risk assessment in the english and welsh correctional services. *Criminal Justice and Behavior* 2012;39:287–307.

⁶ collins Gs, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;13:1.

⁷ eandarme i, Pouls C, De Laender J, et al. Field validity of the HCR-20 in forensic medium security units in Flanders. *Psychology, Crime & Law* 2017;23:305–22.

⁸ Jeandarme i, Edens JF, Habets P, et al. PCL-R field validity in prison and hospital settings. *Law Hum Behav* 2017;41:29–43

⁹ vojtt G, Thomson LDG, Marshall LA. The predictive validity of the HCR-20 following clinical implementation: does it work in practice? *J Forens Psychiatry Psychol* 2013;24:371–85

(également appelés échantillons de dérivation ou de découverte) (figure 1). La transparence des rapports de recherche dans ce domaine est importante, car les conséquences des décisions basées sur certains de ces outils peuvent être préjudiciables à l'individu (sur la base de la sécurité publique).¹⁰¹¹

Validation externe

1. L'outil a-t-il fait l'objet d'une validation externe ? C'est la question la plus importante. Les outils ont tendance à être plus performants dans l'échantillon pour lequel ils ont été créés, comme nous l'expliquons ci-dessous. La seule véritable mesure de la performance d'un outil est sa capacité à prédire le résultat d'un nouvel échantillon distinct¹²¹³

2. Cette validation a-t-elle été effectuée dans la population concernée ? La population de l'échantillon de validation doit présenter des caractéristiques, des facteurs de risque et un risque de base similaires à ceux de l'échantillon de dérivation, et le résultat doit être celui qui nous intéresse. Pour prendre deux exemples, si un outil est requis pour prédire la violence après la sortie de prison, un outil validé pour prédire l'agression chez les patients n'est pas approprié. Ou encore si l'outil n'a été validé que chez les hommes et qu'il est utilisé pour prédire le risque chez les femmes, il surestimerait le risque de récurrence si le sexe n'est pas pris en compte. Dans la pratique, cela expliquerait pourquoi certains outils, tels que la Psychopathy Checklist, qui n'a pas été développée pour prédire le risque de violence, mais pour identifier une forme de trouble de la personnalité, obtiennent des résultats parmi les moins bons des outils couramment utilisés dans une méta-analyse¹⁴. En ce qui concerne l'évaluation du risque de suicide, cela expliquerait également pourquoi de nombreux outils actuellement utilisés ne se sont pas révélés précis dans les études prospectives, car ils ont été développés comme des listes de contrôle des symptômes (par exemple, la Suicide Intent Scale ou SAD PERSONS scale)¹⁵.

3. L'outil repose-t-il sur une méthodologie solide ? Les études de prédiction du risque doivent être basées sur un protocole afin d'éviter la plupart des problèmes décrits ci-dessous. Les études de validation doivent rester fidèles au modèle original et tout changement éventuel doit être spécifié dans le protocole. Dans le cas contraire, l'étude n'est plus une validation externe, mais la dérivation d'un nouveau modèle. La taille de l'échantillon est également cruciale et doit être basée au moins sur 100 événements (résultats) pour la puissance statistique¹⁶ et être un échantillon représentatif pour la généralisation (par exemple, en utilisant des données de cohortes avec un taux élevé de personnes non consentantes qui affectent la validité externe de l'étude, car les personnes qui consentent sont susceptibles d'être différentes de celles qui ne consentent pas par rapport à la prévalence des facteurs de risque d'arrière-plan). Les résultats doivent être publiés dans des revues à comité de lecture, ce qui

¹⁰ Fazel s, Wolf A, Larsson H, et al. Identification of low risk of violent crime in severe mental illness with a clinical prediction tool (Oxford Mental Illness and Violence tool [OxMIV]): a derivation and validation study. *Lancet Psychiatry* 2017;4:461–8

¹¹ Wolf A, Fanshawe TR, Sariaslan A, et al. Prediction of violent crime on discharge from secure psychiatric hospitals: A clinical prediction rule (FoVOx). *Eur Psychiatry* 2017;47:88–93

¹² Monahan J, Skeem JL. Risk assessment in criminal sentencing. *Annu Rev Clin Psychol* 2016;12:489–513.

¹³ Khoury MJ, Gwinn M, Ioannidis JP. The emergence of translational epidemiology: from scientific discovery to population health impact. *Am J Epidemiol*

¹⁴ Singh JP, Grann M, Fazel S. A comparative study of violence risk assessment tools: a systematic review and meta-regression analysis of 68 studies involving 25,980 participants. *Clin Psychol Rev* 2011;31:499–513.

¹⁵ Quinlivan I, Cooper J, Meehan D, et al. Predictive accuracy of risk scales following self-harm: multicentre, prospective cohort study. *Br J Psychiatry* 2017;210:429–36

¹⁶ Collins Gs, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35:214–26.

n'est pas en soi un indicateur de qualité méthodologique, et les publications doivent systématiquement fournir suffisamment de détails méthodologiques pour être reproductibles.

	Good validation study	Good derivation study	Poor derivation
Study type	External validation in new sample	Derivation and internal validation	Derivation only or theoretical tool
Protocol	Protocol based on derivation study	Prespecify analyses, outcomes and so on	No protocol or unclear
Population & outcome	Same as population of interest	Same as population of interest	Different population or outcome
Sample size	>100 events in validation sample	>10 events per variable tested	<10 events per variable tested
Variable inclusion	Same as original tool	Incremental power; face validity	Based on univariate analyses only
Variable weighting	Same as original tool	Based on statistical evidence	Arbitrary weighting
Publication	Published in peer-reviewed journal	Published in peer-reviewed journal	Unpublished or not transparent
Usability	Objective and clearly defined items	Objective and clearly defined items	Subjective rating scales (eg, Likert)
Reporting	Discrimination and calibration measures	Discrimination and calibration measures	No adequate performance measures
Output	Predefined risk levels; absolute risk	Predefined risk levels; absolute risk	Vague or undefined risk categories

Figure 1 Factors to consider in determining the quality of a particular risk assessment tool.

4. L'outil fournit-il des informations essentielles ? Tous les outils doivent présenter à la fois des mesures de discrimination et d'étalonnage. La discrimination mesure la capacité de l'outil à distinguer les personnes atteintes du résultat et ceux qui ne le sont pas, en attribuant une note ou une catégorie de risque plus élevée à ceux qui commettent des infractions. Elle comprend des mesures telles que la sensibilité, la spécificité, la valeur prédictive positive et la valeur prédictive négative, qui peuvent être calculées à des seuils de risque spécifiques, et l'aire sous la courbe, qui est une mesure globale de la discrimination pour tous les seuils de risque possibles. L'étalonnage, quant à lui, consiste à déterminer si l'outil prédit un niveau de risque proche du risque observé. Par exemple, un outil qui prédit un risque d'automutilation de 20 % dans le groupe à haut risque, mais dont seulement 10 % du groupe s'est effectivement automutilé, est mal calibré.

L'étalonnage peut être examiné graphiquement en traçant la courbe du risque prédit par rapport à celle du risque observé, ou par le biais de tests statistiques visant à mesurer le niveau typique d'erreur d'étalonnage.

5. est-il utile, faisable et acceptable ? L'outil doit fournir des informations utiles, y compris des catégories de risques clairement définies, et des risques absolus, le cas échéant. Les éléments doivent également être clairement définis, objectifs et faciles à remplir. Par exemple, les échelles d'évaluation subjectives (par exemple, les échelles de Likert de 1 à 5) seront difficiles à évaluer de manière cohérente et souffriront de variations entre les évaluateurs¹⁷. L'outil doit avoir une validité apparente en incluant des éléments essentiels (par exemple, l'âge et le sexe lorsqu'il s'agit de prédire la violence) et en expliquant l'inclusion, l'importance et la contribution des autres éléments. Il y a des avantages à

¹⁷ Singh JP, Fazel S, Gueorguieva R, et al. Rates of violence in patients classified as high risk by structured risk assessment instruments. *Br J Psychiatry* 2014;204:180–7.

disposer d'outils indépendants de l'entretien pour atténuer le biais de l'observateur, et ces outils ont été recommandés pour l'évaluation du risque de suicide¹⁸.

DERIVATION (ET VALIDATION INTERNE)

Si l'outil n'a pas fait l'objet d'une validation externe, il ne doit pas être utilisé dans la pratique (sauf dans des circonstances inhabituelles, lorsque d'autres solutions ne sont pas appropriées ou disponibles et qu'une validation externe est en cours), en raison d'importantes limitations. Les questions ci-dessous décrivent ces limites comment minimiser leur impact et peuvent être utilisées pour sélectionner les outils candidats appropriés pour une validation externe. Bien entendu, même les outils à validation externe décrits ci-dessus devraient faire l'objet d'une validation prospective après leur mise en œuvre afin de contrôler leurs performances.

1. l'outil suit-il un protocole ? Là encore, il s'agit d'un élément clé si l'on veut que l'étude donne une représentation précise de ses performances. En l'absence de protocole, il est hautement probable que l'on crée un outil qui présente de bonnes mesures de performance, mais dont les résultats sont médiocres dans la pratique. L'échantillon, les variables candidates, les résultats, les périodes de suivi, les analyses statistiques et les résultats doivent tous être préspecifiés avant toute analyse des données.

2. Comment les variables candidates ont-elles été sélectionnées ? Imaginez un outil d'évaluation des risques basé sur des centaines de variables collectées régulièrement. Les analyses révèlent que toute personne née un vendredi présente l'événement en question (par exemple, la criminalité). L'inclusion de cette variable améliore les performances du modèle, mais il s'agit clairement d'une découverte fortuite qui ne sera pas observée dans l'échantillon externe. Plus on teste de variables, plus on trouve d'associations fortuites avec la criminalité et plus la performance du modèle rapportée sera d'autant moins représentative. La règle empirique est que pour chaque variante testée, l'échantillon doit comporter au moins 10 résultats¹⁹. Dans cet exemple, si 20 variables sont testées, l'échantillon doit comporter au moins 200 personnes ayant commis un délit au cours de la période de suivi afin d'éviter l'inclusion d'associations fallacieuses. En outre, l'inclusion des variables doit suivre le protocole et impliquer une forme de régression multivariable pour le déterminer. Sinon, les outils incluront des variables qui n'ajoutent pas de précision prédictive supplémentaire et qui sont plus compliqués et plus longs qu'ils ne devraient l'être.

3. Comment les variables sont-elles pondérées ? De nombreux outils accordent le même poids à tous les éléments inclus. Cela suppose, d'une part, que toutes les variables incluses ont la même association avec le résultat, mais aussi que les variables sont toutes indépendamment liées de manière indépendante au résultat. Le revenu et les antécédents de délinquance sont tous deux associés à un risque plus élevé de délinquance, mais sont-ils aussi importants l'un que l'autre ? Il est très peu probable que ce soit le cas, et les outils qui n'ont pas pondéré les éléments individuels donneront de moins bons résultats²⁰.

4. Comment les autres paramètres ont-ils été sélectionnés ? En l'absence de protocole, et que les chercheurs constatent que leur outil est le plus performant pour un résultat à 6 mois, plutôt qu'à 1, 3

¹⁸ bolton JM, Gunnell D, Turecki G. Suicide risk assessment and intervention in people with mental illness. *BMJ* 2015;351:h4978.

¹⁹ royston P, Sauerbrei W. *Multivariable model - building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables*. Chichester: John Wiley & Sons, 2008.

²⁰ Hamilton Z, Neuilly M-A, Lee S, et al. Isolating modeling effects in offender risk assessment. *J Exp Criminol* 2015;11:299–318.

ou 12 mois, ils pourraient être tentés de l'utiliser comme résultat principal. Cependant, il s'agit là d'une forme de test multiple une forme de test multiple - comme parier sur la victoire de son cheval mais, après l'avoir vu perdre, dire que vous avez en fait parié sur la prochaine course. La conséquence sera que l'outil sera beaucoup moins performant dans le monde réel.

5. Une validation interne a-t-elle été effectuée ? Cette méthode comprend souvent le bootstrapping, qui consiste à prélever un certain nombre d'échantillons aléatoires dans l'ensemble des données afin de fournir une estimation de la précision de la mesure des performances. La division de l'échantillon original en deux groupes aléatoires est une forme de validation interne, car les deux échantillons seront statistiquement équivalents en raison de la randomisation. Cette forme de division est parfois présentée comme une validation externe, mais elle n'est pas due à la distribution égale des variables prédictives. Pour obtenir une validation externe, la division de l'échantillon devrait être basée sur d'autres variables, non liées au résultat²¹.

En résumé, nous avons fourni une liste de contrôle simple en 10 points qui pourrait être appliquée par les cliniciens, les chercheurs et les décideurs politiques pour tester la pertinence d'un outil d'évaluation du risque particulier. Nous recommandons que des recherches soient menées pour déterminer combien d'outils actuellement utilisés répondent à ces critères, et que de nouvelles études prennent en compte ces questions lors de l'élaboration et de la validation de nouveaux instruments dans les domaines de la justice pénale et de la santé mentale.

Article financé par le Wellcome Trust (202836/Z/16/Z).

Intérêts concurrents : Aucun n'a été déclaré.

Provenance et examen par les pairs : Pas de commande ; examen par les pairs en externe.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>
doi:10.1136/eb-2017-102861
Accepted 8 December 2017

²¹ Fazel s, Chang Z, Fanshawe T, et al. Prediction of violent reoffending on release from prison: derivation and external validation of a scalable tool. *Lancet Psychiatry* 2016;3:535–43