

Comparison of artificial intelligence and human-based prediction and stratification of the risk of long-term kidney allograft failure

Divard G, Raynaud M, Tatapudi VS, Abdalla B, Bailly E, Assayag M, Binois Y, Cohen R, Zhang H, Ulloa C, Linhares K, Tedesco HS, Legendre C, Jouven X, Montgomery RA, Lefaucheur C, Aubert O, Loupy A.

Publication Commun Med 2, 150 (2022). <https://doi.org/10.1038/s43856-022-00201-9>

Mots-clés : transplantation – stratification du risque – perte du greffon

Introduction

La capacité à prédire le risque d'un événement d'intérêt en médecine est essentielle à la prise de décision clinique. Actuellement, les cliniciens doivent guider leurs décisions à partir de données de plus en plus nombreuses et complexes. La transplantation rénale illustre bien cette problématique avec de multiples sources de données (cliniques, biologiques, immunologiques et histologiques) parfois contradictoires que le clinicien doit intégrer pour prendre sa décision afin d'optimiser la survie du greffon. Pour accompagner les cliniciens dans leurs décisions, plusieurs modèles de prédictions existent mais aucun n'a encore été comparé à la performance humaine.

L'objectif de cette étude était d'évaluer si ces modèles de prédictions peuvent avoir un intérêt en pratique clinique pour homogénéiser l'évaluation du pronostic des patients en transplantation rénale.

Dans cette étude, c'est l'outil validé iBox récemment qualifié par l'agence européenne du médicament qui a été choisi pour comparer ses performances prédictives à celle de cliniciens en transplantation rénale.

Matériels et méthodes

Il s'agit d'une étude observationnelle incluant neuf cliniciens (3 internes, 3 assistants chefs de clinique et 3 seniors) pour évaluer les dossiers médicaux anonymisés à un an post-transplantation de 400 patients transplantés rénaux entre 2005 et 2014 dans 4 centres de transplantation français. Chaque clinicien devait indiquer un pourcentage de survie à 7 ans post-évaluation (soit 8 ans post-transplantation) pour chacun des 400 patients. Parallèlement, un outil qualifié de prédiction du risque de perte de greffon (le système iBox) était calculé pour

chaque patient avec les mêmes données au moment de l'évaluation à 1 an post-transplantation.

Plusieurs critères d'évaluations ont été investigués :

- 1) La performance pour prédire la survie observée du greffon à 7 ans post-évaluation entre les cliniciens et le système iBox en utilisant la discrimination et la calibration (deux mesures complémentaires pour évaluer la prédiction d'un événement à un temps donné).
- 2) L'accord entre les cliniciens pour prédire la survie à long terme du greffon en utilisant la corrélation intra classe.
- 3) L'Évaluation des paramètres ayant le plus influencés les cliniciens dans leurs prédictions en utilisant un modèle de *random forest* pour chaque clinicien (permet de hiérarchiser les paramètres d'un modèle)

Résultats

A 7 ans post-évaluation, un total de 84 (21%) patients transplantés ont perdu leur greffon.

- 1) La performance prédictive à 7 ans post-évaluation du système iBox était supérieure avec une meilleure discrimination (C- statistique 0.789) et calibration (figure 3) comparée aux cliniciens. En effet, les cliniciens dans cette étude semblaient surestimer le risque de perte de greffon (figure 3).
- 2) Les prédictions des cliniciens étaient hétérogènes avec une distribution des prédictions très variable (figure 2) y compris lorsque les cliniciens avaient le même niveau d'expérience clinique avec une corrélation intra classe respectivement de 0,48 pour les internes, 0,61 pour les assistants et 0,59 pour les seniors.
- 3) Parmi les 10 paramètres ayant le plus influencé les prédictions des cliniciens, le débit de filtration glomérulaire était le plus important et consistant avec un Kappa (Fleiss) de 0,75. En revanche, la hiérarchie des paramètres d'importance était largement différente d'un médecin à l'autre, avec un kappa de Fleiss global de 0,13 (Figure 4) et ce quel que soit l'expérience clinique (table 4 en supplementary). A noter, trois cliniciens n'avaient pas les anticorps spécifiques du donneur parmi les paramètres d'importance.

Conclusion

Dans cette étude, un outil digital démontre de meilleures performances prédictives que neuf cliniciens même expérimentés. Les cliniciens semblent surestimer le risque de perte de greffon et sont très hétérogènes dans leurs prédictions. Cette variabilité pourrait s'expliquer par l'hétérogénéité des paramètres que les cliniciens utilisent pour faire leurs prédictions et ce quelle que soit leur niveau d'expérience clinique.

Ces résultats soutiennent l'utilisation de ce type d'outils digitaux en pratique clinique pour accompagner les cliniciens lors de l'évaluation d'un patient en routine afin

d'homogénéiser les pratiques dans le but d'optimiser la survie des greffons post-transplantation rénale.

Les points forts

- L'une des premières études comparant des cliniciens et un modèle pronostique pour prédire la survenue d'un événement à long terme.
- Question pertinente, souvent posée avec l'arrivée de nouveaux outils digitaux qui prétendent faire mieux que l'humain sans le prouver.
- L'analyse des paramètres ayant le plus influencés les prédictions des cliniciens est un vrai plus pour comprendre l'hétérogénéité entre les cliniciens même lorsque ceux-ci ont le même niveau d'expérience clinique.

Les points faibles

- On regrettera le faible nombre de cliniciens inclus (n=9).
- L'évaluation des patients était à 1 an avec une prédiction à 7 ans uniquement gênant l'extrapolation des résultats en pratique clinique. Un essai randomisé est en cours (ClinicalTrials.gov Identifier: NCT05112315) et compare un suivi avec l'outil iBox versus soin courant pour évaluer la valeur additionnelle en pratique clinique de l'outil.
- Il aurait été aussi intéressant que chaque clinicien donne sa liste de paramètres les plus important pour la prédiction lors de chaque évaluation de patients.

Dr Gillian Divard, premier auteur de la publication