

Proposition de thèse CIFRE en Biostatistique/Epidémiologie

Dans le cadre d'une collaboration entre l'unité INSERM 1246 SPHERE, l'institut de transplantation du CHU de Nantes (ITUN), l'unité INSERM 1064 CRTI, et l'entreprise Sêmeia, nous proposons une thèse CIFRE au sein de l'école doctorale Biologie Santé de l'université Bretagne Loire.

Projet de recherche:

Titre. Développement et validation d'un outil multivarié de prédiction dynamique d'un échec de greffe rénale.

Contexte.

La transplantation rénale est le traitement de référence pour le patient insuffisant rénal en termes de survie, de comorbidités, de qualité de vie et de coût financier pour le système de santé. Cependant malgré une amélioration très significative de la survie des greffons durant la première année de greffe, le pronostic de survie à long terme des greffons reste inchangé depuis 20 ans. Les principaux facteurs de risques précoces et tardifs associés à l'échec de greffe sont bien connus parmi lesquels certains sont associés au greffon et à la greffe et ne seront pas modifiables comme l'âge du donneur ou du receveur (1). En revanche, un certain nombre de facteurs de risque liés au receveur comme les risques cardiovasculaires, métaboliques ou liés aux médicaments immunosuppresseurs pourront être accessibles à une prise en charge médicale appropriée qui pourrait participer à améliorer la survie des greffons et des patients. Enfin, les facteurs de risques immunologiques représentent une part importante de la perte des greffons avec la survenue de rejet. Pour minimiser ce risque immunologique les patients sont soumis à un régime immunosuppresseurs au long cours pour lequel l'adhérence au traitement joue à rôle majeur.

Une revue de la littérature a récemment montré l'existence de scores pronostiques qui pourraient aider à l'amélioration de la prise en charge des patients (2). Cependant, ces scores sont statiques et ne peuvent être calculés qu'à un seul temps post-transplantation. Cette limite pourrait être solutionnée par l'apport des modèles conjoints pour données longitudinales et données de survie qui ont montré leur intérêt pour le développement d'outils de prédictions dynamiques, i.e. prédictions pouvant se mettre à jour au cours du suivi du patient en fonction de l'accumulation de nouvelles informations (3,4).

La créatinémie est un marqueur longitudinal de la fonction rénale bien connu et utilisé en routine dans le suivi du patient transplanté d'un rein. Nous avons récemment démontré son lien avec le risque d'échec de greffe à partir d'un modèle conjoint pour données longitudinales et données de survie (5). Nous avons ainsi récemment développé et validé un outil de prédiction dynamique

(DynPG – Dynamic prediction of Patient and Graft survival) permettant de prédire l'échec de greffe sur une fenêtre de 5 ans (6). Utilisant six variables disponibles à l'inclusion, cet outil peut se mettre à jour en fonction des nouvelles mesures de créatinémie collectées au cours du suivi. Développé et validé en interne à partir de patients transplantés rénaux de la cohorte multicentrique française DIVAT (www.divat.fr), le DynPG vient d'être validé en externe à partir de patients des centres de transplantation de Lille (France) et Louvain (Belgique) (7).

Plusieurs limites du DynPG peuvent être soulignées et suggèrent que des améliorations puissent y être apportées. Le DynPG est notamment développé à partir d'un seul marqueur longitudinal : les mesures répétées de la créatinémie. Les prédictions ainsi obtenues ne peuvent être que le seul reflet des modifications de la créatinémie au cours du temps. Le DynPG n'est alors pas sensible à la survenue d'événements cliniques tels qu'à un cancer ou l'apparition de lésions cérébrovasculaires qui n'auraient pas d'effet sur la créatinémie. Le DynPG doit donc être utilisé avec précautions face à de telles situations cliniques. D'autres marqueurs longitudinaux tels que la protéinurie (8), l'apparition d'une immunisation dirigée contre les antigènes HLA du donneur en post-greffe (9), les variations du taux sanguin de tacrolimus (principal médicament immunosuppresseur post-greffe) (10,11) et l'adhérence au traitement pourraient apporter des informations pour améliorer la prédiction d'échec de greffe au fil du suivi. Une hypothèse de ce projet de recherche est que la non-adhérence au traitement puisse améliorer la prédiction de l'échec de greffe. Nous nous intéresserons notamment au ratio de possession de médicaments (MPR – Medication Possession Ratio) défini comme le nombre effectif de jours de traitement sur une période de temps. Il semble donc important d'intégrer ces marqueurs longitudinaux dans le contexte du pronostic en transplantation rénale.

Objectif.

L'objectif de ce projet est de poursuivre le développement du DynPG. Il s'agira d'étudier l'apport de variables longitudinales mesurées au cours du suivi du patient transplanté rénale dans l'amélioration des capacités prédictives de l'échec de greffe rénale. Nous nous intéresserons plus particulièrement à la non-adhérence au traitement, l'apparition d'une immunisation de novo (dnDSA), d'une protéinurie et d'une lymphopénie au cours du suivi.

Matériel et Méthodes.

Les données utiles à la réalisation de cette étude pourront être extraites de deux bases de données qui seront couplées. D'une part, les données cliniques de patients transplantés seront extraites de la base de données DIVAT.

DIVAT (Données Informatisées et VALidées en Transplantation, www.divat.fr) est une cohorte française multicentrique de patients transplantés rénaux réunissant l'ensemble des paramètres cliniques et biologiques du receveur, du donneur et de la transplantation et les paramètres du suivi de greffe. Les mêmes critères d'inclusion que pour le développement du DynPG seront appliqués : receveurs adultes d'une première ou seconde greffe d'un donneur vivant ou décédé à cœur battant,

et vivant à un an post-greffe maintenus sous tacrolimus et mycophenolate. Les facteurs de risque cliniques du receveur, du donneur et de la transplantation sont disponibles. Les marqueurs longitudinaux d'intérêt seront notamment la protéinurie, l'immunisation anti HLA spécifique du donneur de novo (dnDSA) et la lymphopénie.

D'autre part, afin d'étudier la non-adhérence au traitement et le parcours de soins des patients une demande d'extraction de données auprès du SNDS (Système National des Données de Santé) sera faite pour obtenir des informations sur la délivrance de médicament immunosuppresseurs. Le SNDS est l'évolution du Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM) créé par la loi de financement de la sécurité sociale pour 1999 et mis en œuvre depuis 2003 par la Caisse nationale d'assurance maladie des travailleurs salariés (Cnam). Le SNDS regroupe actuellement les données de l'Assurance Maladie dont les données de consommation inter-régimes et l'échantillon généraliste des bénéficiaires (EGB et EGBS), les données des établissements de santé et les causes médicales de décès. Les données du SNDS sont des observations indirectes du traitement car elles correspondent aux remboursements du parcours de soin des patients comme les transactions en pharmacie, les hospitalisations, etc. Elles comprennent également des informations sur le bénéficiaire. Leur avantage est de contenir l'intégralité de la population française et d'être relativement bien structurée. Leur défaut est l'absence de nombreuses variables cliniques. À partir des données du SNDS, il est possible de définir et étudier différents descripteurs du comportement du patient. Entre autre, il est possible d'estimer le MPR défini comme le ratio du nombre de délivrance sur le nombre de prescription dans une fenêtre de prédéterminé (6 mois) précédant le diagnostic de non-adhérence. Le MPR est une variable quantitative pouvant évoluer au cours du temps traduisant l'intensité de l'adhérence du patient au traitement au cours du suivi. Couplé aux données cliniques de DIVAT, il sera alors envisageable de correctement étudier le rôle de la non-adhérence sur le risque d'échec de greffe en tenant compte d'éventuels facteurs de confusion. De même, d'autres caractéristiques ou descripteurs pourront être définis et évalués à partir des données du SNDS comme par exemple, les différentes pathologies déclarées à l'Assurance Maladie (ALD - affection longue durée), les séjours hospitaliers préalables à la greffe ou intercurrents, les autres traitements prescrits non liés à la greffe, la co-occurrence d'actes de soins (actes infirmiers, consultations...), ou encore le statut social approché par l'éligibilité CMU par exemple.

L'évolution au cours du suivi du patient de ces différents marqueurs répétés sera étudiée à l'aide de modèles statistiques adaptés au caractère longitudinal de ces variables. Afin de comprendre le lien de ces marqueurs avec le risque d'échec de greffe (défini comme le 1er événement entre un retour en dialyse, un décès avec un greffon fonctionnel ou une re-transplantation), nous envisageons le développement d'un modèle conjoint à effets aléatoires partagés pour données longitudinales multivariées et données de survie. Pour un tel développement méthodologique, il faudra en particulier intégrer la corrélation des marqueurs entre eux et la nature des variables qui implique des distributions adaptées.

L'échantillon de patients de la cohorte DIVAT couplé aux données du SNDS sera réparti aléatoirement en une cohorte d'apprentissage et une cohorte de validation. La cohorte d'apprentissage servira à l'estimation du modèle conjoint. Les capacités pronostiques du DynPG multivarié (intégrant plusieurs marqueurs longitudinaux) seront étudiées à partir de la cohorte de validation interne.

Sur un plan méthodologique, des alternatives aux modèles conjoints pourraient s'avérer pertinentes. L'utilisation des modèles de machine learning et d'apprentissage profond pourraient présenter des avantages pour identifier des marqueurs prédictifs de l'échec de greffe. L'un des enjeux de ce projet de recherche porte sur l'intérêt de telles approches comparativement à des modèles plus traditionnelles que sont les modèles conjoints. Nous espérons pouvoir comparer les capacités prédictives d'outils qui seront développés selon les deux approches mais aussi d'évaluer leurs complémentarité. Par exemple, les modèles de deep-learning ont montré des capacités intéressantes pour extraire des caractéristiques de données qui pourraient être ensuite modélisées avec une approche statistique.

Résultats attendus.

Le projet de recherche présenté comprend plusieurs étapes. Un premier travail porte sur le développement d'un modèle conjoint pour données longitudinales et données de survie. Ce travail fera l'objet d'un premier article en épidémiologie de la transplantation rénale. Dans un travail méthodologique, nous comparerons les performances prédictives issus du modèle conjoint à celles de modèles de machine learning. Le modèle retenu permettra de développer un outil multivarié de prédictions dynamiques dont les capacités prédictives seront étudiées selon les recommandations du TRIPOD (12). Cette partie du projet fera l'objet d'un article clinique. En parallèle, nous effectuerons des études en R et Python pour le calcul des prédictions dynamiques en fonction du modèle retenu. Ces programmes Python pourront être intégrés à une librairie développée et exploitée par Sêmeia permettant de facilement calculer ces prédictions dynamiques. Les éléments existants de cette librairie pourront également être utilisés pour ces études.

Au-delà des développements méthodologiques, une importante partie du projet de recherche porte sur le développement d'une solution numérique adaptée au patient et au clinicien pour appréhender le risque d'échec de greffe. Cet outil de prédiction dynamique fera l'objet d'une étude spécifique pour démontrer de son intérêt comme outil d'information au patient et au clinicien. Cette étude ancillaire permettra alors d'envisager le déploiement du DynPG comme Dispositif Médical (DM).

Références.

1. Van Loon E, Bernards J, Van Craenenbroeck AH, Naesens M. The causes of kidney allograft failure: more than allo-immunity. A viewpoint paper. Transplantation 2019;

2. Kaboré R, Haller MC, Harambat J, Heinze G, Leffondré K. Risk prediction models for graft failure in kidney transplantation: a systematic review. *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association* 2017; 32: ii68–ii76.
3. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press, 2012.
4. Proust-Lima C, Blanche P. *Dynamic Predictions*. Wiley StatsRef: Statistics Reference Online 2016;
5. Fournier M-C, Foucher Y, Blanche P, Buron F, Giral M, Dantan E. A joint model for longitudinal and time-to-event data to better assess the specific role of donor and recipient factors on long-term kidney transplantation outcomes. *European Journal of Epidemiology* 2016; 31: 469–479.
6. Fournier M-C, Foucher Y, Blanche P et al. Dynamic predictions of long-term kidney graft failure: an information tool promoting patient-centred care. *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association* 2019; 34: 1961–1969.
7. Lenain R, Dantan E, Giral M et al. External Validation of the DynPG for Kidney Transplant Recipients. *Transplantation* 2020;
8. Foucher Y, Daguin P, Akl A et al. A clinical scoring system highly predictive of long-term kidney graft survival. *Kidney International* 2010; 78: 1288–1294.
9. Loupy A, Aubert O, Orandi BJ et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study. *BMJ (Clinical research ed.)* 2019; 366: l4923.
10. Sapir-Pichhadze R, Wang Y, Famure O, Li Y, Kim SJ. Time-dependent variability in tacrolimus trough blood levels is a risk factor for late kidney transplant failure. *Kidney International* 2014; 85: 1404–1411.
11. Süsal C, Döhler B. Late intra-patient tacrolimus trough level variability as a major problem in kidney transplantation: A Collaborative Transplant Study Report. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2019; 19: 2805–2813.
12. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Annals of Internal Medicine* 2015; 162: 55–63.

Laboratoire d'accueil/Entreprise:

- **INSERM UMR 1246 – SPHERE**

Labélisé en 2017, l'unité INSERM UMR 1246 – SPHERE (methodS for Patient-centered outcomes and HHealth REsearch) est une équipe de recherche multidisciplinaire dont l'objectif est de promouvoir la recherche méthodologique centrée sur le patient et de renforcer la contribution des données rapportées par les patients (qualité de vie, anxiété, bien-être, etc.) dans la prise de décision et l'évaluation des soins. Trois axes de recherche principaux portent sur : 1) la mesure et l'interprétation de l'expérience et la perception du patient, 2) l'évaluation en vie réelle du patient, 3) la médecine de précision centrée sur le patient. Les 3 axes ont pour objectifs de proposer des méthodologies adaptées pour mieux évaluer et comprendre l'adaptation des patients à leur maladie et fournir une meilleure interprétation des résultats rapportés par les patients, pour prendre en compte l'environnement du patient dans l'évaluation pragmatique de la prise en charge du patient et pronostic, pour intégrer les perspectives et les préférences des patients dans la prise de décision médicale et l'évaluation pronostique.

- **INSERM UMR 1064 – CRTI**

Créé en 2012 et renouvelé en 2017, le Centre de Recherche en Transplantation et Immunologie (CRTI INSERM UMR 1064) rassemble des chercheurs et cliniciens spécialisés en immunologie, transplantation, virologie, néphrologie, médecine régénérative, génétique et bioinformatique. Ils développent des projets innovants allant de la recherche fondamentale à la recherche clinique. Les objectifs sont d'améliorer les traitements et le suivi des patients dans le domaine des transplantations et des maladies inflammatoires à médiation immunitaire grâce à la compréhension des réponses immunitaires, au développement de nouvelles immunothérapies, de biomarqueurs et d'outils pour la médecine personnalisée et à la mise au point de stratégies alternatives pour le remplacement d'organes ou de tissu

- **SEMEIA**

Créée en 2017, Sêmeia conçoit des solutions de suivi et d'accompagnement des patients atteints de maladies graves ou chroniques. En utilisant l'Intelligence Artificielle (deep learning et sequence mining), les données des patients notamment issues du Dossier Médical Partagé (DMP) sont croisées avec les données de l'Assurance maladie (SNDS) et permettent de prédire les ruptures dans le parcours de soins (arrêts de traitement, complications, réhospitalisations, etc.). Dans le cadre de ses développements, Sêmeia a développé des algorithmes prédictifs à partir de données du SNDS pour prédire la non-adhérence de patients à un traitement lors d'affections longue durée. L'expertise de Sêmeia a notamment permis de développer des outils dans le contexte du cancer du sein.

Cette thèse se déroulera entre Nantes et Paris et de nombreux déplacements entre ces deux

endroits sont à prévoir. D'autres déplacements permettront également de présenter les travaux et les résultats obtenus.

Encadrement :

- **Etienne Dantan** (MCU en Biostatistique, INSERM UMR 1246 – SPHERE)
Etienne.Dantan@univ-nantes.fr
INSERM UMR 1246 – SPHERE (methodS for Patient-centered outcomes and HEalth REsearch)
IRS2, 22 boulevard Bénoni Goullin, 44200 Nantes
www.sphere-nantes.fr
- **Magali Giral** (PU-PH en Néphrologie, CHU de Nantes - ITUN, INSERM UMR 1064 – CRTI)
Magali.Giral@chu-nantes.fr
Institut de Transplantation, Urologie-Néphrologie (ITUN), CHU Nantes - Hôtel Dieu, 30 Bd Jean Monnet, 44093 Nantes Cedex 01
INSERM UMR 1064 – CRTI (Centre de Recherche en Transplantation et Immunologie)
www.divat.fr
- **Thomas Janssoone** (SEMEIA)
tjanssoone@semeia.io
9 cour des Petites Ecuries, 75010 Paris
www.semeia.io

Profil du candidat et Compétences requises :

- Bonnes connaissances en Biostatistique et Epidémiologie, en particulier des modèles de régression pour l'analyse de données longitudinales et l'analyse de données de survie
- Bonne connaissance des approches de machine learning souhaitable, ainsi que de bibliothèques comme Scikit-learn, Keras, pyTorch, numpy, pandas, ...
- Bonne maîtrise de la programmation Python
- Bonne connaissance des approches de développement et validation d'outils pronostiques
- Goût pour la modélisation de problèmes rencontrés en santé