UroCCR
RÉSEAU FRANÇAIS DE RECHERCHE SUR LE CANCER DU REIN

Fiche de soumission de projet au Comité Scientifique et Ethique

TITRE DU PROJET : Évaluation comparative des méthodes d'analyse statistique sur données synthétiques et réelles en IRM multiparamétrique des petites tumeurs rénales.
Acronyme : **IRM-SYN**
UroCCR n° 255

PROMOTEUR / RESPONSABLE DE TRAITEMENT

☒ ACADEMIQUE ☐ INDUSTRIEL

Responsable(s)/coordonnateur(s) du projet : Pr Jean-Christophe Bernhard
Organisme : CHU de Bordeaux
Adresse : Hôpital Pellegrin, Place Amélie Raba Léon, 33000 BORDEAUX

TYPE DE PROJET

☐ PROSPECTIF ☒ RETROSPECTIF

☐ MONOCENTRIQUE ☒ MULTICENTRIQUE

Date de début des inclusions : 27/11/2018 (1ère inclusion) Période d'étude : 53 mois
Date de fin des inclusions : 14/04/2023 (DVDP)

Partenaires (équipes) associés au projet : ☒ OUI ☐ NON

Si oui, compléter les informations suivantes pour chacun des partenaires :

☒ Intra-UE ☐ Hors-UE

Responsable(s) : Ségolène Combettes, Sylvain Francoise Dal-Mas
Organisme : Alia Santé
Raison sociale : ALIA SANTE
Rôle du partenaire : Industriel / Génération de données synthétiques

DESCRIPTION / JUSTIFICATION SCIENTIFIQUE DU PROJET

Contexte :
Le cancer du rein, classé au 14e rang des cancers les plus fréquents dans le monde, connaît une incidence en constante augmentation, avec 431 288 nouveaux cas recensés en 2020 [1]. À ce jour, le traitement de référence des petites masses rénales repose sur la néphrectomie partielle. Pourtant, environ 20 % de ces tumeurs s'avèrent bénignes après exérèse. Le développement de techniques préopératoires non invasives, capables de caractériser précisément la nature et l'agressivité tumorale, est donc essentiel pour optimiser la stratégie thérapeutique. L'imagerie par résonance magnétique multiparamétrique (IRMmp), en intégrant plusieurs séquences d'imagerie fonctionnelle et anatomique, s'impose comme une méthode prometteuse, fiable et reproductible pour distinguer les tumeurs bénignes des tumeurs malignes [2–4]. Dans ce contexte, le projet IRMK01 (UroCCR n°20) a été mené au sein du réseau français du cancer du rein UroCCR, pour évaluer la performance diagnostique de l'IRMmp (test index) en comparaison avec l'histologie (test de référence). Cependant, bien que performante, cette technique reste coûteuse et chronophage, ce qui allonge considérablement la durée des essais cliniques.

Face à ces contraintes, l'utilisation des données synthétiques émerge comme une solution innovante pour accélérer les processus de recherche, tout en assurant un haut niveau de qualité et de sécurité des données médicales. Plusieurs études ont en effet démontré que l'intégration des patients synthétiques dans le cadre des essais cliniques montre des avantages économiques, éthiques, logistiques [5–7].
Le projet IRM-SYN, porté conjointement par le réseau médico-scientifique national UroCCR et Alia Santé, pionnier européen dans la génération de données synthétiques, s'inscrit dans cette dynamique. Il a pour objectif d'exploiter une partie des données de vie réelle issues de l'étude IRMK01 pour produire des données synthétiques, qui permettront d'élargir les cohortes disponibles tout en garantissant la confidentialité des données sensibles. Cette approche ouvre ainsi la voie à une optimisation des processus de gestion des données, à une accélération des essais cliniques et, dans le cas spécifique, à une amélioration du diagnostic de malignité des tumeurs rénales.

Objectif Principal :
Vérifier que les données synthétiques sont représentatives des données réelles.

Objectifs Secondaires :
- Évaluer si l'intégration de données synthétiques permet de renforcer la significativité statistique des analyses.
- Optimiser, grâce aux données synthétiques, les résultats de l'étude IRMK01 afin de développer un modèle de classification capable de déterminer la malignité tumorale, et intégrer un index permettant le calcul d'un score de sévérité.

Critères de jugement :
Le critère de jugement principal sera la capacité des données synthétiques à reproduire fidèlement les propriétés statistiques et cliniques des données réelles d'IRM multiparamétriques du rein. Cela inclura la comparaison de distributions, de corrélations et d'indicateurs descriptifs entre bases réelles et synthétiques.

Les critères de jugement secondaires porteront sur :
- La puissance statistique des analyses : évaluer si l'ajout de données synthétiques permet de renforcer la significativité des résultats des tests menés dans l'étude clinique.
- La cohérence des tendances : vérifier si les analyses conduites uniquement sur données synthétiques aboutissent aux mêmes orientations et conclusions que celles réalisées sur données réelles.
- Les performances du modèle de classification : mesurer, sur un jeu de test composé exclusivement de données réelles, la capacité du modèle entraîné avec données réelles enrichies de données synthétiques à prédire la malignité tumorale.

Hypothèse :
La donnée synthétique permet de faciliter le diagnostic de malignité tumorale basé sur IRM multiparamétrique.

Critères d'inclusion :
Mêmes critères d'inclusion que l'étude IRMK01 :
Age ≥ 18 ans
Tumeurs rénales solides non héréditaires.
Indication de chirurgie rénale ou de biopsie rénale pour suspicion de tumeur maligne
Masse rénale unique
Masse rénale de taille comprise entre 1,5 et 4 cm.
Découverte fortuite par technique d'imagerie.
Consentements éclairés IRMK01 et UroCCR signés.
Patient affilié ou bénéficiaire du régime de sécurité sociale
Toutes les femmes en capacité de procréer doivent être sous contraception efficace jusqu'à la réalisation de l'IRM.

Critères d'exclusion :
Mêmes critères d'exclusion de l'étude IRMK01 :
Signes de malignité patents de la tumeur rénale (métastases, adénopathie, thrombus...).
Lésions kystiques selon la classification de Bosniak.
Lésions présentant de la graisse macroscopiquement visible en échographie ou scanner.
Masses rénales multiples ou bilatérales.
Preuves histologiques disponibles à l'inclusion.
Antécédents de néoplasie rénale quel que soit la localisation ou contexte familial (Von Hippel Lindau, Sclérose de Bourneville).
Insuffisance rénale modérée à terminale documentée (clairance de la créatinine <30 ml/min selon MDRD ou CKD-EPI).
IRM non réalisable :
Pacemaker cardiaque (spécialement les anciens modèles),
Pompes à insuline,
Aide(s) auditive(s),
Neurostimulateur(s),
Clip(s) métallique(s) intracrânien(s),
Eclat(s) métallique(s) intraoculaire(s),
Contre-indication(s) à l'injection de sels de gadolinium.
Refus de l'intervention chirurgicale, de la biopsie, par le patient, si elle est nécessaire.
Mineur.
Personne privée de liberté.
Personne sous tutelle.
Personne sous curatelle.
Personne sous sauvegarde de justice.
Femme enceinte ou allaitante.
Femme en âge de procréer sans test de grossesse négatif avant l'examen IRM.
Adulte en incapacité d'exprimer son consentement.
Suivi clinique impossible pour des raisons psychologiques, familiales, sociales ou géographiques.

Analyses statistiques :
Les analyses statistiques viseront dans un premier temps à comparer de manière rigoureuse les distributions, corrélations et caractéristiques descriptives des données synthétiques par rapport aux données réelles, afin de vérifier leur représentativité. Des métriques quantitatives (distances statistiques, mesures de similarité multivariée ...) seront combinées à des analyses visuelles (réduction de dimensions, histogrammes comparatifs ...) pour évaluer la fidélité des données générées. En parallèle, le degré d'anonymisation sera aussi évalué afin de garantir le caractère confidentiel des données générées.

Dans un second temps, nous évaluerons, sur l'étude clinique elle-même, l'impact de l'intégration des données synthétiques sur la puissance statistique des tests. L'objectif est de déterminer si leur ajout permet de renforcer la significativité des résultats, et si les analyses menées sur les données synthétiques seules conduisent à des tendances et conclusions similaires à celles obtenues avec les données réelles.

Enfin, les données enrichies permettront de développer et d'optimiser un modèle de classification destiné à prédire la malignité tumorale sur la base des IRMmp, en intégrant un index de calcul d'un score de sévérité. L'entraînement sera réalisé sur un jeu mixte (réel + synthétique) mais l'évaluation finale sera conduite exclusivement sur des données réelles, afin de garantir la validité clinique et éviter tout biais lié à l'utilisation de données synthétiques en phase de test.

Résultats attendus :
Améliorer la précision diagnostique en intégrant des données synthétiques aux analyses existantes.

Bibliographie :
1. Bukavina L, Bensalah K, Bray F, et al. Epidemiology of Renal Cell Carcinoma: 2022 Update. Eur Urol. 2022;82(5):529-542. doi:10.1016/j.eururo.2022.08.019
2. Sasiwimonphan K, Takahashi N, Leibovich BC, Carter RE, Atwell TD, Kawashima A. Small (<4 cm) Renal Mass: Differentiation of Angiomyolipoma without Visible Fat from Renal Cell Carcinoma Utilizing MR Imaging. Radiology. 2012;263(1):160-168. doi:10.1148/radiol.12111205
3. Cornelis F, Tricaud E, Lasserre AS, et al. Routinely performed multiparametric magnetic resonance imaging helps to differentiate common subtypes of renal tumours. Eur Radiol. 2014;24(5):1068-1080. doi:10.1007/s00330-014-3107-z
4. Cornelis F, Lasserre AS, Tourdias T, et al. Combined Late Gadolinium-Enhanced and Double-Echo Chemical-Shift MRI Help to Differentiate Renal Oncocytomas With High Central T2 Signal Intensity From Renal Cell Carcinomas. Am J Roentgenol. 2013;200(4):830-838. doi:10.2214/AJR.12.9122
5. Gonzales A, Guruswamy G, Smith SR. Synthetic data in health care: A narrative review. Johnson A, ed. PLOS Digit Health. 2023;2(1):e0000082. doi:10.1371/journal.pdig.0000082

6. Lai TC, Ngorsuraches S. Roles of AI-Based Synthetic Data in Health Economics and Outcomes Research. Value Health. Published online April 2025:S1098301525023125. doi:10.1016/j.jval.2025.04.2157
7. Foraker R, Morrow JD, Johnson JA, Wilcox AB, Forster AJ, Payne PRO. Understanding synthetic data: artificial datasets for real-world evidence. BMJ Evid-Based Med. Published online July 2, 2025:bmjebm-2024-113617. doi:10.1136/bmjebm-2024-113617

DONNEES CLINIQUES / RESSOURCES BIOLOGIQUES
Afin de faciliter la sélection des données, merci de répondre à ces questions :

- Quelles sont les données UroCCR que vous souhaitez mobiliser pour ce projet ?

☒ Données démographiques : âge, sexe, phénotype, …

☒ Données de diagnostic : date, IMC, ASA, ECOG, antécédents, cTNM, …

☒ Données d'imagerie loco-régionale : date, type, description tumorale, scores RENAL et PADUA, …

☒ Données d'imagerie d'extension : dates, types d'imagerie, résultats, adénopathies et métastases synchrones (localisations et nombre de sites), …

☒ Données biologiques : date, NFS (Hb, Leucocytes, PNN, Lymphocytes, plaquettes), dosages biochimiques et enzymatiques (créatininémie, GGT, ASAT, ALAT, PAL, LDH, protéinémie, albuminémie, Ca, Ca corrigé, CRP, clairance (MDRD, Cockroft, CKD-EPI), …

☒ Données de biopsie : date, côté, type histologique, grade de Fuhrman/ISUP, …

☒ Données de chirurgie : date, ASA, côté, description détaillée de la procédure (type de chirurgie, voie d'abord, clampage, sutures, durée opératoire, …), complications per-op et post-op, reprise chirurgicale, transfusion, durée d'hospitalisation, créatininémie, Hb, clairance (MDRD, Cockroft, CKD-EPI), …

☒ Données d'anatomopathologie : date, côté, type histologique, taille tumorale, description détaillée de la pièce opératoire (envahissements, architectures, composantes cellulaires), pTNM, grade de Fuhrman/ISUP, marges chirurgicales, scores pronostics (UISS et SSIGN), …

☐ Données de ressources biologiques : date, type d'échantillon, numéro d'identification, poids ou volume de l'échantillon, disponibilité de l'échantillon, temps avant cryoconservation.

☒ Données de traitements ablatifs : date, ASA, côté, description détaillée de la procédure (type, technique de détection et d'accès, applicateurs, durée opératoire), complications per-op et post-op, transfusion, durée hospitalisation, créatininémie et clairance (MDRD, Cockroft, CKD-EPI), …

☒ Données de traitement oncologique : Indication, date d'initiation et d'arrêt, nombre de cycles, scores pronostiques (MSKCC/HENG), molécule (DCI, dose et unité), pause, interruption, changement de dose, évaluation tumorale (critères RECIST v1.1), toxicités (type, grade 3 à 5, imputabilité au traitement), …

☒ Données de suivi : date, évaluation clinique (poids, ECOG, Karnofsky, hémodialyse), nouvelles complications, nouvelles comorbidités, récidive, progression, décision thérapeutique, statut vital, …

☐ Données oncogénétiques : prélèvement réalisé, tests effectués, résultats des tests, …

☐ Données de radiothérapie : dates, côté, nombre de fractions, dose totale, asservissement respiratoire, isodose de prescription, indice de compromission, GTV, PTV, …

☐ Données de PROMs et PREMs : Questionnaires RAAC, EQ-5D-5L, e-Satis, Wake, STAI-Trait, STAI-Etat, HLSEU-Q16, SF-12, HADS, …

- Recours données chaînées au Système National des Données de Santé (SNDS) ☐ OUI ☒ NON

- Le projet a-t-il des besoins en ressources biologiques ? ☐ OUI ☒ NON

*** Type :**
*** Quantité :**
*** Autres précisions :**

CALENDRIER PREVISIONNEL

Evaluation du projet par le Comité : 08/09/2025

Extraction des données UroCCR : à définir

Génération des données synthétiques : à définir

Analyses statistiques : à définir

Soumission abstract : à définir

Soumission article : à définir

Revue ciblée : à définir

Réglementation :
La mise en œuvre des projets au sein de l'Entrepôt de Données de Santé (EDS) UroCCR se fait en accord avec les principes du règlement (UE) 2016/679 du Parlement Européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (dit RGPD), à la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (dite LIL), au référentiel relatif aux traitements de données à caractère personnel mis en œuvre à des fins de création d'entrepôts de données dans le domaine de la santé de la CNIL (dit référentiel EDS) et de charte de fonctionnement du réseau UroCCR en vigueur au jour de la soumission de la présente Fiche Projet. L'ensemble des données issues de la base UroCCR sont confidentielles et ne doivent donc pas être communiquées aux personnes non autorisées à les traiter. Par ailleurs, le traitement des données pour une autre finalité que celle décrite dans la présente Fiche Projet n'est pas autorisé.

Cadre réservé à l'équipe coordinatrice UroCCR (ou UroCCR-Chain si données chainées)

Faisabilité et Evaluation scientifique et éthique

Disponibilité des données

Date : à définir Oui : ☒ Non : ☐
Effectifs prévisionnels : 335 patients

Disponibilité des ressources biologiques dans la biocollection virtuelle

Date : Oui : ☐ Non : ☐ NA : ☒
Commentaires :

AVIS DU COMITÉ SCIENTIFIQUE ET ETHIQUE
Date : 08/09/2025
Evaluation du projet :
Positif ☒ Négatif ☐ En attente ☐
Commentaires :