

Université de Montréal

Développement et validation interne de scores de risque clinique afin de prédire le risque de tumeurs rénales et pulmonaires chez les gens atteints de la sclérose tubéreuse de Bourneville

*Par*

Frédéric Loubert

Département de médecine sociale et préventive

École de santé publique de l'Université de Montréal

Mémoire présenté en vue de l'obtention du grade de Maîtrise ès science (M.Sc.)

en Épidémiologie

Mai 2023

© Frédéric Loubert, 2023

Université de Montréal

Département de médecine sociale et préventive

École de santé publique

---

*Ce mémoire intitulé*

**Développement et validation interne de scores de risque clinique afin de prédire le risque de tumeurs rénales et pulmonaires chez les gens atteints de la sclérose tubéreuse de Bourneville**

*Présenté par*

**Frédéric Loubert**

*A été évalué(e) par un jury composé des personnes suivantes*

**Helen Trottier**  
Président-rapporteur

**Mark R. Keezer**  
Directeur de recherche

**Nathalie Auger**  
Membre du jury



## Résumé

**Contexte :** La sclérose tubéreuse de Bourneville (STB) affecte approximativement une naissance vivante sur 6000. Les principales sources de morbidité chez les enfants sont des manifestations cliniques affectant le cerveau et le cœur. Quant aux adolescents et aux adultes, ils sont à plus haut risque de développer des angiomyolipomes rénaux (AML) et la lymphangioliomyomatose pulmonaire (LAM).

**Objectif :** Ce projet de recherche a pour principal objectif de déterminer la capacité du fardeau de la maladie chez l'enfant à prédire le développement futur d'AML et de LAM chez l'adolescent et l'adulte.

**Méthodes :** Les données sur 2420 participants de la TSC (Tuberous sclerosis complex) Alliance Natural History Database ont été utilisées pour développer des modèles de régression logistique afin de prédire les AML et la LAM. Ces modèles comptaient comme variables dix manifestations cliniques de la STB en plus du sexe biologique et de la mutation génétique. Les modèles développés ont été convertis en scores de risque clinique et une validation interne a ensuite été effectuée avec rééchantillonnage par bootstrap afin d'évaluer leur performance prédictive à l'aide de mesures de discrimination et de calibration.

**Résultats :** 1000 participants ont été diagnostiqués avec un AML, alors que la LAM a été détectée chez 107 participants. Les scores de risque clinique les moins élevés prédisaient des risques d'AML et de LAM de 0 % et de 5 %, tandis que les scores les plus élevés pouvaient prédire des risques allant jusqu'à 96 % et 54 %, respectivement. La performance prédictive des modèles allait de bonne à excellente (C-Index variant de 0,75 à 0,85).

**Conclusion :** Les scores de risque clinique développés indiquent une bonne capacité des manifestations cliniques sélectionnées à prédire le risque futur des tumeurs, principalement pour les AML. Une validation externe devra cependant être menée afin de confirmer les résultats obtenus dans la présente étude.

**Mots-clés :** STB, angiomyolipomes, lymphangioliomyomatose, prédiction

## Abstract

**Background:** Tuberous sclerosis complex (TSC) affects approximately 1/6000 live births. The primary sources of childhood morbidity and mortality are manifestations affecting the brain and heart. Adolescents and adults, on the other hand, are at the greatest risk of developing renal angiomyolipoma (AML) and pulmonary lymphangiomyomatosis (LAM).

**Objective:** The main objective of this research project is to determine the ability of the burden of TSC in children to predict the risk of developing AML and LAM in adults.

**Methods:** Data from 2420 participants in the TSC Natural History Database were used to develop logistic regression models to predict AMLs and LAM. Variables used for the models included 10 clinical manifestations of TSC, in addition to biological sex and genetic mutation. The models were converted into clinical risk scores and were internally validated using bootstrap resampling to evaluate their predictive performance with the help of discrimination and calibration measures.

**Results:** 1000 participants were diagnosed with AML, while LAM was seen in 107 participants. The lowest clinical risk scores predicted a risk of AML and LAM of 5% and 0%, while the highest scores predicted a risk of 96% and 54%, respectively. The predictive performance of the models ranged from good to excellent (C-Indexes of 0.75 to 0.85).

**Conclusions:** The clinical risk scores indicated a good ability for the selected clinical manifestations to predict the future risk of tumours, particularly of AML. External validation of the risk scores will be important to confirm the robustness of our findings.

**Keywords:** TSC, angiomyolipoma, lymphangiomyomatosis, prediction

# Table des matières

Chapitre 1 — Introduction .....	1
Chapitre 2 — Recension des écrits .....	2
2.1. Sclérose tubéreuse de Bourneville .....	2
2.1.1. Définition et épidémiologie descriptive.....	2
2.1.2. Diagnostic .....	3
2.1.3. Manifestations cliniques .....	4
2.1.4. Traitements .....	5
2.2. Angiomyolipomes rénaux.....	6
2.2.1. Définition et épidémiologie descriptive.....	6
2.2.2. Diagnostic et suivi.....	7
2.2.3. Complications .....	8
2.2.4. Traitements .....	9
2.3. Lymphangioléiomyomatose pulmonaire .....	9
2.3.1. Définition et épidémiologie descriptive.....	9
2.3.2. Diagnostic et suivi.....	10
2.3.3. Complications .....	11
2.3.4. Traitements .....	11
2.4. Modèles de prédiction.....	12
2.4.1. Définition .....	12
2.4.2. Développement .....	13
2.4.3. Validation et évaluation de la performance .....	15
2.4.4. Présentation des résultats .....	17
2.4.4. Études à visée prédictive versus études à visée étiologique .....	20

2.4.5. Prédiction associée au développement des AML et des LAM .....	21
2.5. Retour sur la littérature .....	23
Chapitre 3 — Objectif.....	25
Chapitre 4 — Méthodes .....	26
4.1. Devis et contexte de l'étude.....	26
4.2. Population de l'étude .....	26
4.3. Variables .....	27
4.3.1. Prédicteurs.....	27
4.3.2. Issues d'intérêt .....	30
4.3.3. Critères d'exclusion .....	31
4.5. Analyses statistiques .....	32
4.5.1. Préparation des données.....	32
4.5.2. Description des caractéristiques des participants.....	35
4.5.3. Analyses principales .....	35
4.5.4. Validation et évaluation de la performance .....	38
4.5.5. Présentation des résultats .....	42
4.5.6. Analyses de sensibilité.....	44
4.6. Considérations éthiques .....	45
Chapitre 5 — Résultats .....	46
5.1. Article .....	46
5.2. Tableaux.....	60
5.3. Figures.....	65
5.4. Suppléments en ligne .....	67
5.5. Contribution des auteurs .....	74
Chapitre 6 — Discussion .....	75

6.1. Sommaire des résultats .....	75
6.2. Forces de l'étude .....	78
6.3. Considérations méthodologiques .....	79
6.3.1. Validité externe .....	79
6.3.2. Validité interne.....	80
6.4. Contributions et implications cliniques .....	85
6.5. Orientations futures.....	87
6.6. Conclusion .....	88
Références bibliographiques .....	89
Annexes.....	93
Annexe 1. Mini-protocole.....	93
Annexe 2. Commentaires de Simon De Montigny suivant la présentation du mini-protocole .....	103
Annexe 3. Commentaires de Nathalie Auger suivant la présentation du mini-protocole.....	104
Annexe 4. Certificat d'éthique émis par le Comité d'éthique de la recherche du CHUM .....	105



## Liste des tableaux

Tableau 1. –	Critères de diagnostic de la STB .....	3
Tableau 2. –	Prédicteurs considérés .....	28
Tableau 3. –	Issues d'intérêt.....	30

# Liste des figures

Figure 1. – Exemple fictif de nomogramme..... 19

## Liste des sigles et abréviations

AML: Angiomyolipome

IRM : Imagerie par résonance magnétique

LAM : Lymphangioliomyomatose

SEGA : Subependymal giant cell astrocytoma

STB : Sclérose tubéreuse de Bourneville

TAND : TSC-associated neuropsychiatric disorder

TSC: Tuberous sclerosis complex

TSC-NHD : Tuberous sclerosis complex Natural history database

## **Remerciements**

J'aimerais dans un premier temps remercier mon directeur de recherche, Mark Keezer, pour son support constant tout au long de mon parcours à la maîtrise. Un merci particulier pour tes nombreux conseils, ta disponibilité, ainsi que pour le partage de ton expertise qui m'ont permis de pouvoir réaliser ce mémoire.

Je souhaite également remercier tous mes proches pour leur soutien et leurs encouragements qui m'ont été d'une grande aide à chacune des étapes allant de mon admission à la maîtrise au dépôt de ce mémoire.

# Chapitre 1 — Introduction

La STB est une maladie génétique qui affecte approximativement une naissance vivante sur 6000 [1]. Elle est une des causes génétiques primaires de l'épilepsie. La principale source de morbidité vue chez les enfants provient de manifestations survenant au niveau du cerveau, du cœur ainsi que de la peau. Quant aux adolescents et aux adultes, ils sont à plus haut risque de développer des angiomyolipomes rénaux et des lymphangioliomyomatoses pulmonaires [1].

La STB atteste une grande hétérogénéité phénotypique, pouvant être observée chez des membres d'une même famille qui présentent la même mutation génétique [2]. Le fardeau associé à la maladie est aussi largement affecté par l'âge de l'individu.

Jusqu'à maintenant, les études publiées se sont limitées à montrer que l'âge, le sexe biologique féminin et les mutations génétiques du gène TSC2 comparativement au gène TSC1 sont associés à un risque plus élevé d'AML rénal et de LAM pulmonaire [3-5]. La connaissance de ces facteurs de risque ne suffit toutefois pas à adapter les recommandations en matière de dépistage pour les patients atteints de la STB. Ils sont donc contraints à passer des tests d'imagerie en série jusqu'à la fin de leurs vies [6].

Il n'existe aucune donnée présentement dans la littérature visant à déterminer si des caractéristiques plus discriminatives basées sur les manifestations cliniques de la maladie durant l'enfance permettraient d'identifier les adolescents ou les adultes à bas ou à haut risque de contracter l'une des deux complications. Ainsi, nous avons décidé de nous pencher sur cette question et d'étudier l'impact que pourraient avoir ces manifestations de la STB sur le risque de développement des AML rénaux et des LAM pulmonaires.

## Chapitre 2 — Recension des écrits

La recension des écrits débute avec un portrait de la maladie d'intérêt, la sclérose tubéreuse de Bourneville, et se poursuit avec les deux issues, les AML et les LAM. S'en suivra une revue de la littérature portant sur le développement de modèles de prédiction cliniques dans laquelle seront détaillées les méthodes actuellement utilisées et les distinctions entre ce type d'étude et les autres études dans le domaine de l'épidémiologie. Le tout se terminera avec un retour sommaire sur les connaissances existantes en matière de prédiction du développement des AML et des LAM chez les personnes atteintes de la sclérose tubéreuse de Bourneville.

### 2.1. Sclérose tubéreuse de Bourneville

#### 2.1.1. Définition et épidémiologie descriptive

La sclérose tubéreuse de Bourneville est une maladie dominante autosomale caractérisée par le développement de tumeurs bénignes dans de multiples systèmes d'organes [1]. Elle est causée par une mutation qui survient soit sur le gène *TSC1* ou le gène *TSC2*.

On estime que la prévalence à la naissance de la STB se place entre une sur 6000 et une sur 10 000 naissances vivantes [1]. La prévalence de la maladie ne parait pas être influencée par le sexe d'une personne, quoique celui-ci semble toutefois avoir un impact sur la présence et la sévérité de certaines manifestations [7]. Une mutation du gène *TSC2* est en cause de la maladie dans la majorité des cas avec une prévalence se situant autour de 70 %, comparativement à la mutation *TSC1* qu'on aperçoit dans 20 % des cas [1]. Les 10 % restants comprennent les gens chez

qui aucune mutation n'a pu être identifiée, potentiellement en raison de la présence de mosaïcisme somatique [1].

## 2.1.2. Diagnostic

L'*International Tuberous Sclerosis Complex Consensus Group* a mis à jour en 2021 les critères de diagnostic de la sclérose tubéreuse qui avaient été élaborés en 2013. D'après ceux-ci, un diagnostic peut être obtenu en raison de la présence de manifestations cliniques de la maladie ou bien suivant les résultats d'un test génétique. [8, 9].

Les critères utilisés pour établir le diagnostic clinique sont répertoriés comme étant soit majeurs ou bien mineurs et sont détaillés au Tableau 1. Un diagnostic est certain si au minimum deux critères majeurs ou un critère majeur combiné à deux critères mineurs sont détectés. Un diagnostic est possible s'il y a présence d'un critère majeur seul ou d'au moins deux critères mineurs. Un diagnostic définitif peut aussi être obtenu indépendamment des manifestations cliniques si une mutation pathogène des gènes *TSC1* ou *TSC2* est identifiée suivant un test génétique sur des tissus normaux [8].

Tableau 1. – Critères de diagnostic de la STB

Critères majeurs	Critères mineurs
Macules hypopigmentées ( $\geq 3$ ; diamètre d'au moins 5 mm)	Lésions cutanées « Confetti »
Angiofibromes ( $\geq 3$ ) ou plaque fibreuse céphalique	Puits de l'émail dentaire ( $\geq 3$ )
Fibromes unguéaux ( $\geq 2$ )	Fibromes intra-oraux ( $\geq 2$ )
Peau de chagrin	Tache achromique rétinienne
Hamartomes rétiniens multiples	Kystes rénaux multiples
Dysplasies corticales ( $\geq 3$ )	Hamartomes extra-rénaux
Nodules sous-épendymaires ( $\geq 2$ )	Lésions osseuses sclérosées
Tumeur sous-épendymaire à cellule géante	

---

Rhabdomyome cardiaque

---

Lymphangioliéiomyomatose

---

Angiomyolipomes ( $\geq 2$ )

---

### **2.1.3. Manifestations cliniques**

Les manifestations cliniques associées à la sclérose tubéreuse sont multiples et varient grandement d'un individu à l'autre, tant en nombre qu'en sévérité. [1]. Les AML et les LAM, deux des principales complications rénales et pulmonaires de la maladie ne seront pas abordées dans cette partie puisqu'elles seront explorées plus en détail dans les deux sections du chapitre suivant celle-ci.

Parmi les manifestations les plus communes de la maladie, on retrouve celles affectant la peau. Présentes chez la majorité des individus atteints de STB, elles comprennent entre autres les angiofibromes faciaux qui sont des lésions s'apparentant à de l'acné, les macules hypopigmentées, des taches pâles bénignes pouvant apparaître à n'importe quel endroit sur le corps et la peau de chagrin, une région de peau surélevée habituellement située dans le bas du dos [1]. Ces manifestations surviennent en règle générale durant l'enfance et ne sont que rarement la cause de complications graves [8].

Le cerveau est aussi dans la plupart des cas touché par la sclérose tubéreuse. L'épilepsie constitue le symptôme le plus observé auprès des personnes avec la STB, se trouvant chez 80 % à 90 % d'entre eux [1]. Les troubles neuropsychiatriques associés à la STB sont aperçus chez près des deux tiers des individus et leur présentation peut varier énormément d'une personne à l'autre. Parmi ceux-ci, on retrouve un niveau de déficience intellectuelle pouvant aller de faible à sévère, de l'autisme, de l'anxiété et de la dépression. L'ensemble des troubles neuropsychiatriques



associés à la STB sont regroupés sous le terme des *TSC-associated neuropsychiatric disorders* (TAND) [8]. Les nodules sous-épendymaires et les tumeurs sous-épendymaires à cellules géantes sont des masses présentes sur le cerveau dans 80 % à 90 % et dans 10 % à 15 % des cas respectivement et sont une grande source de morbidité rattachée à la maladie [1].

La plus répandue des manifestations concernant le cœur est la présence de rhabdomyomes cardiaques. Ces tumeurs bénignes fréquemment découvertes avant la naissance lors de l'échographie n'occasionnent que rarement des complications et tendent à régresser durant l'enfance. Leur prévalence passe de 90 % pour les enfants en bas âge à 20 % pour les adultes [1].

Les manifestations se présentant sur les reins sont communes et peuvent engendrer de graves complications. Outre les AML qui sont explorés plus en détail dans la prochaine section, on retrouve les kystes rénaux et les carcinomes à cellules rénales. Ces dernières sont des tumeurs malignes dont la prévalence est estimée entre 2 % et 3 % chez les personnes atteintes de la STB. On les associe néanmoins à un important risque de mortalité, ayant occasionné deux des 16 décès observés dans une étude qui portait sur les causes de mortalité des individus vivant avec la STB [10].

En somme, ces manifestations ne sont que quelques-unes des principales manifestations touchant les individus atteints de la sclérose tubéreuse. D'autres organes peuvent en être affectés, tels qu'entre autres les yeux et les dents, quoiqu'avec un impact moindre sur le fardeau de la maladie [1].

#### **2.1.4. Traitements**

La sclérose tubéreuse étant une maladie qui demeure à ce jour incurable, les traitements qu'un individu recevra viseront spécifiquement à guérir ou à réduire le fardeau associé aux

manifestations de celle-ci. Étant donné qu'il n'existe que peu de moyens de prédire la survenue des différentes complications, leur gestion reposera sur un suivi actif dans l'optique d'observer la progression des symptômes [8].

Parmi les plus récentes avancées dans le traitement de manifestations cliniques de la sclérose tubéreuse, on retrouve la prise de médicaments inhibiteurs de mTOR. Trois essais cliniques sont arrivés à des résultats prometteurs concernant le traitement à l'aide d'un inhibiteur de mTOR, l'everolimus, pour trois conditions représentant une grande source de morbidité pour les gens avec la STB. Le premier essai, EXIST-1, a associé la prise d'everolimus à une importante baisse dans la taille des tumeurs sous-épendymaires à cellules géantes SEGA [11]. Le second, EXIST-2, a trouvé que le médicament diminuait le volume des angiomyolipomes rénaux [12]. Le dernier, EXIST-3, a montré l'efficacité de l'everolimus en tant que traitement d'appoint pour la réduction de la fréquence des crises épileptiques [13]. Un grand nombre d'études motivées par les résultats des trois essais cliniques sont encore menées à ce jour afin d'explorer la capacité des inhibiteurs de mTOR de guérir et même de prévenir dans certains cas d'autres manifestations cliniques dont souffrent les gens vivant avec la STB [14].

## **2.2. Angiomyolipomes rénaux**

### **2.2.1. Définition et épidémiologie descriptive**

Les angiomyolipomes rénaux sont définis comme des tumeurs bénignes composées de cellules graisseuses, de vaisseaux sanguins et de muscles provoquant des lésions sur les reins [3]. Bien qu'on les associe généralement à la STB, elles se développent dans la majorité des cas de façon sporadique dans la population générale [15]. Cependant, les manifestations des AML diffèrent considérablement selon que la sclérose tubéreuse en soit la cause ou non. On retrouve

entre autres une fréquence plus élevée de tumeurs bilatérales et multiples ainsi qu'un risque de croissance des lésions supérieur lorsque la sclérose tubéreuse en est la cause [15]. Les cas sporadiques se produisent en général à un âge plus tardif, soit autour de 40 à 50 ans [15]. Étant donné ces différences au niveau des manifestations et les objectifs de ce projet de recherche, cette section concernera les cas dans lesquels ils sont associés à un diagnostic de STB.

Les AML sont caractérisés en fonction de leur présence sur les deux reins ou sur un seul (bilatéraux ou unilatéraux) ainsi que selon le nombre et le diamètre des lésions qu'ils contiennent [7]. Ces tumeurs sont communes parmi les personnes atteintes de la STB avec une prévalence se situant autour de 47 % et pouvant augmenter jusqu'à plus de 70 % pour les individus de plus de 18 ans [7]. Bien que l'âge médian du diagnostic des AML soit évalué à approximativement 13 ans [3], certains cas peuvent être identifiés avant même la première année de vie, tandis que d'autres ne le sont que plus tard à l'âge adulte [7]. Les AML sont présents chez les hommes et les femmes, avec une prévalence légèrement plus élevée au sein de ces dernières [7].

### **2.2.2. Diagnostic et suivi**

On peut distinguer trois méthodes permettant de détecter la présence d'un AML rénal ; l'IRM, la tomodensitométrie et l'échographie [16]. L'IRM est en règle générale proposée comme modalité de choix, du fait de sa capacité supérieure aux autres approches de repérer les AML. L'IRM s'avère particulièrement efficace lorsque les tumeurs comprennent peu de cellules graisseuses, un phénomène se produisant dans 25 % à 30 % des cas [8]. Malgré cela, son coût élevé, son accessibilité réduite dans la pratique clinique ainsi que la difficulté pour certains individus avec la STB de rester immobiles pour toute la durée de l'examen n'en font pas une option viable dans toutes les situations. On suggère en second lieu la tomodensitométrie, qui tout comme l'IRM présente d'excellentes sensibilité et spécificité pour le dépistage de masses sur les reins

[15]. À cela s'ajoutent un faible coût, une grande disponibilité dans la pratique clinique et un examen plus rapide. La tomодensitométrie n'est toutefois pas toujours indiquée pour les personnes atteintes de la STB qui doivent fréquemment avoir recours aux imageries, dues aux conséquences associées à l'exposition répétée aux rayonnements ionisants [15]. L'échographie demeure une option sécuritaire, peu coûteuse et facilement accessible. Cependant, on le suggère moins pour un premier diagnostic en raison de son incapacité à détecter la présence d'anévrismes et de masses contenant peu de cellules graisseuses, ces dernières étant plus communes chez les enfants et dans les tumeurs récentes [8].

Les recommandations actuelles en matière de dépistage de l'AML proposent qu'une imagerie abdominale soit planifiée dès qu'un diagnostic de STB est obtenu. Un suivi devrait ensuite être fait tous les 1 à 3 ans si l'examen initial n'a pas permis de repérer de tumeurs. À partir du moment où un AML est découvert, le suivi passe aux 6 à 12 mois, dans l'optique d'observer de près la progression des tumeurs [8]. Ce suivi se doit d'être maintenu tout au long de la vie, des lésions pouvant faire leur apparition ou bien commencer à croître indépendamment de l'âge de la personne atteinte [1].

### **2.2.3. Complications**

Bien qu'on estime à plus de 80 % la proportion des cas ne présentant aucun signe ou symptôme au moment de leur diagnostic [7], les complications que les AML peuvent engendrer sont fréquentes et peuvent avoir un impact considérable sur la santé. Effectivement, les complications rénales sont la première cause de décès chez les personnes atteintes de sclérose tubéreuse [10]. La principale complication associée aux tumeurs est le développement de saignements sur les reins dont le risque d'apparition se situerait entre 9 % et 21 % [1]. Ces

saignements peuvent rapidement mener à des hémorragies internes spontanées s'ils ne sont pas traités à temps [16].

#### **2.2.4. Traitements**

Les traitements pour les AML varient selon la présence de symptômes et le diamètre des lésions existantes sur les tumeurs. Ceux-ci sont idéalement prophylactiques, ayant pour objectif de prévenir la survenue de complications, de réduire la taille des lésions ou bien de ralentir leur progression [1]. Le traitement de première intention recommandé pour un individu présentant des tumeurs asymptomatiques, en accroissement, et montrant des lésions d'un diamètre de plus de trois centimètres est la prise d'inhibiteurs de mTOR (everolimus). Ce médicament qui a montré des résultats prometteurs dans le cadre d'un essai clinique de phase 3 permet de diminuer la croissance des AML et même de la renverser dans certains cas [12]. Mis à part son efficacité, ce traitement est privilégié puisqu'il n'occasionne que peu d'effets secondaires et n'a pas de répercussions sur les fonctions rénales. Néanmoins, lorsqu'il y a présence de saignements, un traitement chirurgical est recommandé. On favorisera l'embolisation des vaisseaux sanguins à une résection chirurgicale si possible, afin de réduire le risque de complications engendrées par une perte de fonction rénale [16].

### **2.3. Lymphangioliomyomatose pulmonaire**

#### **2.3.1. Définition et épidémiologie descriptive**

La lymphangioliomyomatose est la principale manifestation pulmonaire de la STB. Elle est caractérisée par l'apparition de tumeurs composées de cellules musculaires lisses dans les poumons, l'accumulation de celles-ci menant à la formation de kystes pulmonaires [1]. Tout comme pour l'AML, la LAM peut se développer de manière sporadique ou bien être due à la

présence de la sclérose tubéreuse. La distinction entre les deux variantes de la maladie se veut pourtant moins claire étant donné la rareté de la LAM sporadique, sa prévalence se trouvant entre 3,4 et 7,8 cas par millions de femmes dans la population générale [17].

La LAM affecte presque exclusivement les femmes adultes. Dans la cohorte TOSCA [7], 94 % des cas répertoriés de LAM étaient des femmes et parmi celles-ci, 98,7 % étaient âgées de plus de 18 ans.

La prévalence de ces tumeurs pulmonaires chez les femmes avec STB varie selon les publications, se situant la plupart du temps entre 26 % et 49 % [4]. Toutefois, certaines études considèrent que la prévalence serait nettement supérieure, pouvant aller jusqu'à 80 % chez les femmes d'âge adulte [7]. En effet, l'âge semble avoir un impact majeur sur la prévalence de la maladie. Une revue rétrospective des résultats d'imageries par tomodensitométrie faites sur des femmes âgées de plus de 15 ans atteintes de STB sur une période de 12 ans a permis d'estimer la prévalence de la LAM par groupe d'âge. Celle-ci était de 27 % pour les 21 ans et moins, de 24 % pour les 21 à 29 ans, de 58 % pour les 30 à 40 ans et de 81 % pour les 40 ans et plus [4].

### **2.3.2. Diagnostic et suivi**

Le diagnostic initial de la LAM se fait à l'aide d'une tomodensitométrie à haute résolution, souvent accompagnée d'un test des fonctions pulmonaires [4]. Il est actuellement recommandé pour toutes les femmes de passer ces examens dès l'âge de 18 ans, qu'il y ait ou non présence de signes ou de symptômes. L'imagerie devrait par la suite être répétée à tous les 5 à 7 ans dans l'absence de symptômes, tandis que les fonctions pulmonaires devraient être vérifiées plus régulièrement, jusqu'à une fois par année [8]. Si l'examen a permis de détecter la présence de

LAM, la fréquence du suivi variera selon plusieurs critères, tels que les complications accompagnant la LAM et les traitements reçus jusqu'à présent [8].

### **2.3.3. Complications**

Les estimations portant sur la proportion des cas étant symptomatiques varient selon les études, passant de 40,3 % à 63 % [4, 7]. La dyspnée, une sensation de difficulté respiratoire ou d'essoufflement, est le symptôme le plus fréquent et souvent le premier aperçu chez ceux qui présentent des complications [7, 16]. Étant une maladie chronique et incurable, la sévérité des complications y étant reliées tend à augmenter au fil des années. Parmi les plus sévères, on retrouve le pneumothorax et l'affaissement des poumons, les deux contribuant à la diminution graduelle des capacités pulmonaires [7]. La gravité des complications n'est pas négligeable, portant la LAM à la troisième position en importance pour les causes de mortalité associées à la STB, tout juste derrière les lésions rénales et du cerveau [16]. La LAM a d'ailleurs mené au décès de 12,5 % des personnes qui en étaient atteintes dans l'étude citée précédemment [4].

### **2.3.4. Traitements**

Puisque la lymphangioléiomyomatose ne peut être guérie, les traitements visent essentiellement à stabiliser la condition et à en prévenir la progression. Les inhibiteurs de mTOR, plus particulièrement le sirolimus, constituent le traitement de première intention [8]. Ce médicament s'est montré efficace dans la stabilisation des fonctions pulmonaires et dans la diminution des symptômes chez certains patients dans l'essai clinique de phase 2 MILES [18]. On recommande aussi la prise de bronchodilatateurs afin d'aider à l'ouverture des voies respiratoires, ce qui permet la réduction des symptômes de dyspnée [7]. En derniers recours, la transplantation

pulmonaire demeure une option de traitement viable pour la LAM en phase terminale. La survie médiane suivant cette procédure a d'ailleurs été estimée à 12 ans [19].

## **2.4. Modèles de prédiction**

### **2.4.1. Définition**

Les modèles de prédiction sont des outils visant à déterminer la probabilité qu'un événement spécifique survienne en se basant sur la combinaison de multiples prédicteurs [20, 21]. Lorsqu'utilisés dans un contexte clinique, on les divise en deux catégories avec des ambitions distinctes. On retrouve de prime abord les modèles de prédiction à visée diagnostique. Leur objectif est généralement de déterminer le risque qu'une maladie soit actuellement présente chez un individu [22]. La deuxième catégorie est celle des modèles de prédiction à visée pronostique. Dans cette dernière, l'issue qu'on tente de prédire ne s'est toujours pas produite, le but du modèle étant de déterminer son risque de développement futur [20, 21].

Étant donné les objectifs de ce projet de recherche, cette section du chapitre portera essentiellement sur les études à visée pronostique. Les modèles pronostiques trouvent leur utilité tant dans la recherche que dans la pratique clinique. En recherche, leur application principale réside dans la sélection de participants dans les études expérimentales, permettant aux chercheurs d'identifier avec facilité des personnes étant à haut risque de développer l'issue étudiée [21]. Dans la pratique clinique, les utilisations des modèles de prédiction validés sont diverses. Ils peuvent entre autres servir à informer les patients de leur risque de développer une maladie, de tomber en rémission, de complications et de décès. On les emploie aussi afin d'évaluer la probabilité qu'un traitement se révèle efficace [21]. Ces connaissances permettent d'aider les médecins et leurs



patients à déterminer conjointement les suivis et les traitements les plus appropriés en fonction du risque individuel que l'issue représente pour le patient [21].

## 2.4.2. Développement

Le processus menant à l'implémentation d'un modèle de prédiction dans la pratique clinique courante comprend plusieurs étapes, en commençant par le développement du modèle. Plusieurs aspects doivent être considérés à cette étape tels que le choix du devis d'étude, la taille de l'échantillon et la sélection des prédicteurs [23].

De multiples devis d'étude sont utilisés lors du développement de modèles de prédiction, tous avec leurs avantages et leurs inconvénients. Diverses sources, incluant les directives TRIPOD (*Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis*), précisent qu'une cohorte longitudinale avec suivi prospectif créée spécifiquement dans le cadre du développement d'un modèle de prédiction est le devis idéal [21, 24]. Les avantages comprennent entre autres un meilleur contrôle de la mesure des prédicteurs et de l'issue, minimisant ainsi la perte au suivi et les valeurs manquantes [24]. Malgré cela, ils ne sont que peu aperçus dans la littérature actuelle, notamment en raison des coûts élevés et de la durée reliés au suivi prospectif. La majorité des études publiées exploitent des jeux de données déjà existants, provenant de cohortes qui peuvent être rétrospectives ou prospectives, étant donné la simplicité associée à leur utilisation. Une attention particulière doit alors être portée aux valeurs manquantes pour les prédicteurs ainsi qu'aux pertes au suivi chez les participants [24].

En ce qui a trait à la taille d'échantillon, il n'existe pas d'approche universellement adoptée permettant d'estimer avec certitude la quantité de participants requise lors du développement d'un modèle [24]. Cependant, le nombre d'événements de l'issue par variable guide les

recommandations en matière de taille d'échantillon [21, 24]. Cette mesure consiste à diviser le nombre de fois que l'issue étudiée survient par le nombre de prédicteurs présents dans le modèle. Une règle approximative encore majoritairement appliquée à ce jour a établi que les données utilisées devraient contenir au moins dix événements par variable. Ce seuil minimal, fixé selon deux études empiriques [25, 26], permettrait l'obtention d'estimations stables. Depuis, nombre de publications ont remis en question ces recommandations. On avance d'un côté qu'il est dans la plupart des cas possible d'obtenir des estimations adéquates avec un nombre d'événements par variable inférieur à dix [27], tandis que de l'autre côté, on suggère des valeurs supérieures allant même jusqu'à 40 événements par variable [28]. Cependant, puisque dans la pratique courante les chercheurs n'ont que rarement le contrôle sur la taille des échantillons auxquels ils ont accès, la mesure du nombre d'événements par variable n'est souvent utilisée qu'à des fins descriptives [24].

La première sélection de variables à inclure dans le développement d'un modèle repose en général sur des connaissances cliniques de l'issue permettant d'identifier les variables potentiellement pertinentes [21]. Celles-ci peuvent entre autres comporter de l'information sociodémographique sur les participants, les signes et symptômes d'une maladie, les résultats de tests de diagnostic et de dépistage ainsi que les traitements reçus. Afin de permettre l'implémentation future d'un modèle dans la pratique clinique, on recommande que les prédicteurs soient bien définis et qu'ils puissent être reproduits [21]. Un modèle issu de cette première sélection de variables choisies sur la base de connaissances cliniques et non sur la signification statistique des prédicteurs se nomme habituellement le modèle complet [24]. Les chercheurs peuvent cependant être menés à vouloir réduire la quantité de prédicteurs à intégrer dans leurs modèles, en raison par exemple d'un nombre trop élevé rendant difficile son usage dans la pratique clinique [23, 24]. À cet effet, de multiples méthodes de sélection sont applicables. Le pas-à-pas

par élimination en amont (*backward stepwise*) est un procédé automatisé de sélection des variables. Elle débute à partir d'un modèle incluant tous les prédicteurs potentiels et les retire un à la fois selon un critère préétabli tel que l'AIC ou la valeur-p jusqu'à l'obtention d'un modèle final [29]. Malgré sa simplicité d'utilisation, cette technique comporte de nombreux inconvénients, comprenant entre autres une surestimation des coefficients de régression, un risque accru d'overfitting, d'optimisme ainsi qu'une capacité de prédiction réduite [23, 24, 30]. Une alternative recommandée afin de remédier aux désavantages de la précédente approche est le LASSO (*least absolute shrinkage and selection operator*), une méthode de contraction des coefficients de régression [30, 31]. Cette contraction pénalise les coefficients de la régression, menant certains à atteindre une valeur de zéro et de les exclure du modèle final [31]. Bien que cette méthode n'engendre pas les mêmes inconvénients que la précédente, il est important de mentionner qu'elle ne tend pas à améliorer la performance prédictive et qu'elle ne doit donc pas être appliquée à cet effet. L'objectif principal visant son utilisation est habituellement l'obtention d'un modèle simplifié pour lequel la perte de performance prédictive comparativement au modèle complet est minimisée.

### **2.4.3. Validation et évaluation de la performance**

La validation d'un modèle de prédiction et l'évaluation de sa performance sont deux étapes cruciales dans le processus menant à son utilisation dans la pratique clinique. La validation nous assure que le modèle qui a été développé performe de manière satisfaisante lorsqu'il est confronté à de nouvelles données tandis que l'évaluation de la performance nous renseigne quant à la capacité qu'a notre modèle de bien prédire l'issue [23].

Il est recommandé que toute étude de développement de modèle de prédiction comprenne au minimum une certaine forme de validation interne [20]. Ce type de validation qui implique de

n'utiliser que l'échantillon d'origine peut se faire de plusieurs façons différentes. Les données peuvent être préalablement divisées en deux parties, l'une pour le développement et l'autre pour la validation, comme c'est le cas avec la validation par partition d'échantillon (*split-sample*). Cette méthode est cependant portée à surévaluer les capacités prédictives d'un modèle, étant donné la similarité entre les deux jeux de données [32]. Une approche fréquemment utilisée est la validation par rééchantillonnage par bootstrap, qui consiste à créer un nombre prédéfini d'échantillons par rééchantillonnage avec remise à partir d'un seul échantillon initial. Cette approche tend à produire des estimations ayant une plus grande stabilité et à réduire l'optimisme dans la mesure des performances [28].

Bien que la validation interne parvienne à fournir un aperçu de l'efficacité d'un modèle confronté à de nouvelles données, elle ne suffit pas à assurer sa généralisabilité [32]. Il est alors recommandé qu'une validation externe soit effectuée avant de pouvoir considérer l'implémentation d'un modèle de prédiction dans la pratique clinique [20, 21]. Cette forme de validation est la plupart du temps conduite dans une autre étude, parfois par les mêmes chercheurs ayant développé le modèle, parfois par une seconde équipe [20]. Afin de pouvoir mener une validation externe dans un nouvel échantillon, celui-ci doit contenir des données sur toutes les variables incluses dans le modèle qu'on tente de valider [32]. Ce dernier sera testé et les résultats seront comparés avec l'issue réellement observée dans les données. Les mesures de performance seront alors comparées avec celles du modèle initial, dans l'espoir que la diminution en performance prédictive généralement constatée dans cette deuxième cohorte soit minime et que le modèle demeure valide [24].

L'évaluation de la performance d'un modèle de prédiction repose sur deux concepts, soit la discrimination et la calibration. La discrimination correspond à la capacité qu'a un modèle de

distinguer entre ceux qui développeront l'issue et ceux qui ne la développeront pas. On utilise la plupart du temps le c-index qui équivaut à l'aire sous la courbe ROC dans un modèle de régression logistique en vue d'en évaluer la capacité discriminative [24]. Il consiste à comparer des paires d'individus, l'un qui développera l'issue et l'autre qui ne la développera pas afin de vérifier la fraction dans laquelle le modèle attribue une plus grande probabilité de développer l'issue à ce premier [23]. L'obtention de cette valeur nous informe sur la capacité qu'a notre modèle de distinguer entre les individus présentant un risque élevé de l'issue et ceux pour qui le risque est faible. Elle se veut particulièrement importante pour cette raison dans le contexte de l'utilisation d'un modèle en clinique. La calibration correspond à l'écart entre les issues observées et les issues prédites par le modèle. On l'illustre généralement à l'aide d'une courbe de calibration où plus l'accord entre les proportions d'issues et les probabilités prédites est élevé, plus la courbe se rapprochera d'une pente de 45 degrés [23]. Bien que la mesure de la calibration d'un modèle de prédiction soit elle aussi importante, étant donné qu'elle compare les risques moyens entre différents groupes plutôt que le risque individuel, une attention particulière est par conséquent portée à l'évaluation de la discrimination lorsque l'on vise l'utilisation d'un modèle dans un contexte clinique. On recommande que des méthodes de discrimination et de calibration soient appliquées au cours des validations internes et externes d'un modèle afin de vérifier si la performance apparente se rapproche de la performance ajustée dans l'échantillon de validation [32].

#### **2.4.4. Présentation des résultats**

De prime abord, il est recommandé d'inclure dans la présentation des résultats l'équation du modèle complet afin de faciliter la mise en place d'une validation externe par une équipe indépendante [20]. De nombreuses approches permettent par la suite de présenter les résultats d'un

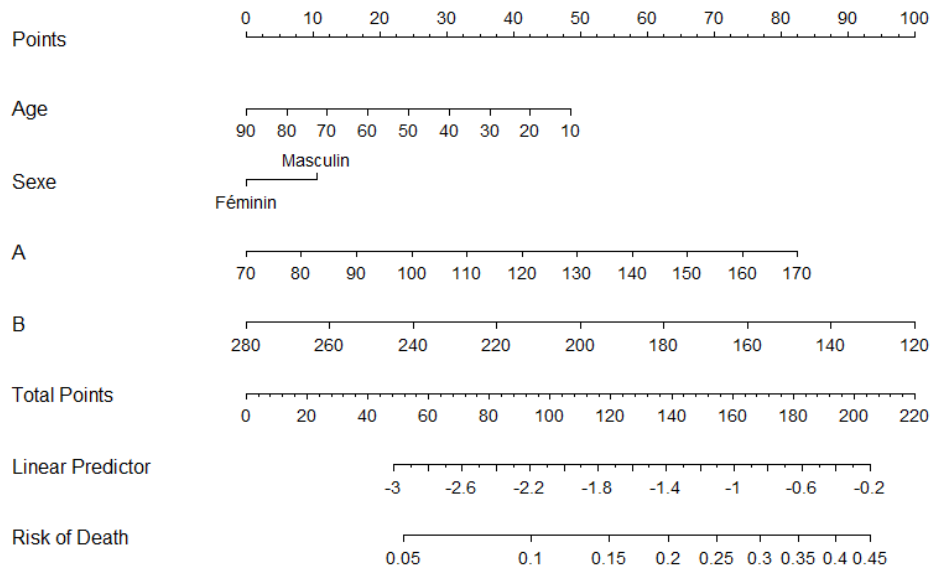
modèle de prédiction et bien qu'ils comportent tous leur lot d'avantages et de désavantages, il n'existe pas de suggestion générale quant à la technique à utiliser. Le choix optimal découle de facteurs liés au développement du modèle ainsi que du contexte dans lequel il sera exploité [33]. Deux des méthodes les plus fréquemment employées seront décrites dans les prochains paragraphes.

La première des deux approches de présentation des résultats est le score de risque clinique. Celle-ci consiste à arrondir les coefficients obtenus à la suite de la régression d'après un facteur commun, attribuant ainsi un pointage à chaque prédicteur. La probabilité de survenue de l'issue sera ensuite mesurée en fonction de la somme des points pour un individu [33]. Elle peut être illustrée de différentes façons, par exemple en utilisant deux tableaux. Le premier tableau montre le score associé à chaque prédicteur et le second la probabilité de survenue de l'issue selon le score total. Le plus important avantage de cette méthode est sa facilité d'utilisation. En effet, les médecins et leurs patients peuvent aisément calculer le risque pour ce dernier en faisant la somme des points reliés aux prédicteurs. Son principal désavantage réside dans le fait que cette méthode n'est qu'une approximation du modèle original. Les variables continues doivent obligatoirement y être catégorisées et tous les coefficients sont arrondis. Cette imprécision peut mener à une perte de capacité prédictive pour le modèle. Il convient donc que la performance des modèles de score de risque soit réévaluée afin de s'assurer qu'elle ne s'éloigne pas de la performance du modèle initial [33].

La seconde approche consiste à illustrer les résultats à l'aide d'un nomogramme, une représentation graphique des résultats d'un modèle de prédiction. Les nomogrammes sont généralement conçus avec l'aide de logiciels statistiques tels que R ou STATA. Un exemple fictif de nomogramme se trouve à la Figure 1. On retrouve dans le haut les points allant de 0 à 100 qui

seront attribués à chacun des prédicteurs. Le positionnement des valeurs pour chaque prédicteur indique le pointage auquel chacun est associé lorsqu'une ligne verticale est dressée jusqu'au score au haut de la figure. Le score total ainsi que la probabilité de survenue de l'issue se trouvent au bas du nomogramme. Pour chaque prédicteur, on doit dresser une ligne verticale reliant sa valeur au score entre 0 et 100, ce qui permet de calculer la somme des points pour chacun. Cette somme correspond au score total. Finalement, afin d'obtenir la probabilité de survenue de l'issue, il suffit de relier le score total à la probabilité associée au bas de la figure [34]. Le principal avantage de la méthode est que l'on peut utiliser le nomogramme dans un contexte où l'accès aux ordinateurs est restreint. Son emploi ne requiert qu'une copie du nomogramme sur papier. Ils peuvent toutefois être complexes à utiliser pour une personne non initiée et devenir difficilement interprétables lorsqu'un modèle contient un grand nombre de prédicteurs [24, 33].

Figure 1. – Exemple fictif de nomogramme



#### **2.4.4. Études à visée prédictive versus études à visée étiologique**

En recherche épidémiologique, les études à visée étiologique et celles à visée prédictive représentent deux domaines séparés, mais souvent confondus. Néanmoins, des distinctions claires et sans équivoques les divisent tant au niveau des objectifs que sur le plan de l'interprétation de leurs résultats [35].

La première distinction réside au niveau des objectifs qui motivent la poursuite d'une étude. La recherche étiologique a pour cible principale l'acquisition d'une meilleure compréhension d'une issue en observant la relation potentiellement causale l'unissant à un facteur de risque spécifique [35]. Quant aux études prédictives, elles visent plutôt à obtenir la meilleure prédiction possible du risque de développer une issue, sans égard au rapport causal pouvant exister entre celle-ci et les autres variables [21, 35].

L'usage de modèles de régression multiple est répandu dans les deux types d'études. Dans les études étiologiques, son emploi sert principalement à contrôler pour la présence de confusion alors que l'on considère uniquement les effets d'un seul facteur de risque d'intérêt. L'identification des variables de confusion y est un concept primordial permettant une estimation non biaisée de l'association [35]. En prédiction, le modèle est plutôt vu comme un entier, où chacune des variables introduites sera interprétée et contribuera à la prédiction [24]. La notion de confusion ne s'applique pas puisque le choix des variables à inclure ne repose pas sur la causalité, mais plutôt sur divers aspects comprenant la performance prédictive et la simplicité d'utilisation du modèle dans un contexte de prédiction clinique [32].

Les différences au niveau des objectifs et des méthodes entraînent une répercussion sur l'interprétation des résultats des modèles dans chacun des types d'étude. En recherche étiologique,



un risque relatif est calculé. Celui-ci nous informe à propos de l'association entre le facteur de risque et la survenue de l'issue. En recherche prédictive, aucune interprétation causale des résultats ne devrait être faite. Les résultats qui sont présentés sous la forme de risques absolus pour chaque variable indépendante constituent plutôt l'impact de chacune d'entre elles sur la probabilité de développer l'issue [35].

Finalement, les concepts de validation et d'évaluation de la performance qui guident les études de prédiction et déterminent la valeur d'un modèle et son potentiel à être employé dans la pratique clinique ne sont pas utilisés dans les études à visée étiologique. Ces études ne ciblant pas la prédiction de l'issue, la qualité des prédictions n'a pas à y être évaluée [35].

#### **2.4.5. Prédiction associée au développement des AML et des LAM**

Tout d'abord, il n'existe à notre connaissance aucune étude publiée à ce jour dans la littérature ayant comme objectif spécifique la prédiction du risque de développement d'AML ou de LAM. Cependant, plusieurs études descriptives sont arrivées à des conclusions similaires concernant certains facteurs de risque qui semblent rattachés non seulement au développement, mais également à la sévérité des tumeurs, tant pulmonaires que rénales.

De nombreuses publications montrent un lien entre les mutations sur le gène *TSC2* et la sévérité de la maladie, l'associant à une plus importante morbidité que les mutations sur le gène *TSC1* [5, 36]. Une série de cas provenant des dossiers cliniques et radiographiques de 167 patients avec la sclérose tubéreuse a établi que les AML étaient plus répandus, plus souvent multiples et bilatéraux et menaient à un plus grand nombre de complications sévères nécessitant une intervention lorsqu'un patient présentait une mutation sur *TSC2* [3]. Une autre série de cas, celle-ci portant sur les résultats d'imageries abdominales et de tests génétiques faits chez des personnes

avec la STB cherchait à évaluer la relation entre la mutation génétique et la LAM [37]. Ils en sont arrivés au constat que bien que la prévalence de la condition ne parût pas différer selon la mutation génétique, *TSC2* serait associé à une quantité plus élevée de kystes sur les poumons, et donc à une manifestation plus sévère de la LAM.

Le sexe biologique d'un individu avec la sclérose tubéreuse semble aussi avoir un impact sur la manifestation de la maladie au cours de sa vie. Certaines manifestations cliniques varient tant dans leur fréquence que dans leur sévérité selon le sexe de la personne affectée [1]. L'étude citée auparavant s'est penchée sur le sujet et a conclu que les AML étaient plus répandus chez les patients de sexe biologique féminin, et ce même lorsque contrôlés pour l'âge et la mutation génétique et que le risque de tumeurs multiples et bilatérales était lui aussi plus élevé. Toujours est-il que, à l'inverse à la mutation génétique, aucun lien n'a été établi entre le risque de complications sévères et le sexe biologique [3]. Une autre étude décrivant la cohorte TOSCA comprenant 2093 participants a relevé que des 987 cas rapportés d'AML, 42,5 % étaient des hommes, tandis que 57,5 % étaient des femmes, montrant encore une fois le lien entre la prévalence des tumeurs rénales et le sexe [7]. La relation entre le sexe biologique et le développement de tumeurs pulmonaires est bien connue, cette manifestation de la STB survenant presque uniquement chez les femmes. Plusieurs hypothèses subsistent pour justifier la présence quasi exclusive chez les femmes pour cette condition, bien qu'aucune ne l'explique totalement. Les LAM sont présentes en majorité chez les femmes préménopausées et une aggravation des symptômes a été observée lors de la grossesse et de la prise de médicaments contenant de l'estrogène. Ces aspects ont mené à l'hypothèse selon laquelle le développement et la progression de LAM seraient dépendants des hormones sexuelles féminines [1]. Bien que certains cas soient

détectés chez des hommes [37], ceux-ci ne sont que très peu associés à une diminution des fonctions pulmonaires [1, 38].

Enfin, l'âge est un facteur important à prendre en considération au niveau des manifestations de la STB. Les principales sources de morbidité et de mortalité associées à la maladie divergent grandement entre l'enfance et l'âge adulte [1]. Une corrélation positive a été montrée entre l'âge d'une personne et la prévalence des AML, indépendamment des effets du sexe et de la mutation génétique [3]. Dans la cohorte TOSCA, la prévalence des AML augmentait à chaque catégorie d'âge, passant de moins de 30 % chez les 5 ans et moins à plus de 70 % pour les 18 ans et plus [7]. Les tumeurs pulmonaires surviennent dans la quasi-totalité des cas chez des individus d'âge adulte, avec une progression plus rapide pour les femmes n'ayant pas atteint la ménopause [1, 7].

## **2.5. Retour sur la littérature**

La recherche dans le domaine de la sclérose tubéreuse a évolué considérablement au cours des dernières années. D'importantes avancées ont d'ailleurs pu être observées au niveau des traitements prophylactiques, surtout avec les inhibiteurs de mTOR. De plus, les recommandations liées au suivi et au traitement de la maladie reposent sur des bases solides qui sont fréquemment mises à jour. Il n'en demeure pas moins que la maladie a encore un impact considérable sur la santé de ceux en étant atteints. Cet impact se fait en particulier sentir lorsque l'on constate la prévalence élevée des tumeurs rénales et pulmonaires et les complications sévères qu'elles peuvent engendrer.

À cet effet, un écart subsiste dans les connaissances liées à la prédiction du développement des AML et des LAM. Cette recension des écrits ne nous a pas permis de repérer dans la littérature

une seule étude de prédiction telle que décrite dans ce chapitre se penchant sur l'identification de prédicteurs du développement des tumeurs. La majorité des publications identifiées suivaient des cohortes de petites tailles et se limitaient à des analyses descriptives comparant la fréquence des AML et des LAM selon certains facteurs. Elles ont tout de même permis d'établir que le risque serait supérieur pour les femmes, qu'il serait corrélé positivement avec l'âge et qu'il paraît fortement associé à la mutation génétique *TSC2* plutôt qu'à la mutation *TSC1*. La connaissance de ces trois facteurs de risque demeure depuis plusieurs années et aucun autre prédicteur du développement des tumeurs ne semble avoir été découvert depuis.

La capacité des manifestations cliniques de la STB survenant durant l'enfance à prédire le risque de développement des tumeurs rénales et pulmonaires n'est à ce jour pas connue et n'a, à notre connaissance, toujours pas été étudiée. Pour cette raison, et pour les implications possibles que sa compréhension pourrait avoir sur le suivi et le traitement des AML et des LAM, cette question mérite d'être approfondie à partir d'une large cohorte et selon les principes de la recherche se rapportant au développement de modèles pronostiques.

## **Chapitre 3 — Objectif**

L'objectif de cette étude était de développer et d'effectuer la validation interne d'indices de risque clinique permettant de prédire l'incidence d'angiomyolipomes rénaux et de lymphangioléiomyomatoses pulmonaires chez les personnes atteintes de la sclérose tubéreuse de Bourneville à partir de manifestations cliniques de la maladie se produisant durant l'enfance.

## **Chapitre 4 — Méthodes**

### **4.1. Devis et contexte de l'étude**

Ceci est une étude de développement et de validation interne de modèles de prédiction utilisant une cohorte longitudinale avec collecte de données ambispective (issue d'une combinaison de méthodes à la fois prospectives et rétrospectives).

### **4.2. Population de l'étude**

Ce projet de recherche s'est appuyé sur des données provenant de la TSC-NHD, une base de données toujours active qui est maintenue par la TSC Alliance depuis son instauration en 2006 [39]. Son financement est assuré en entier par la TSC Alliance sur la base de dons venant de diverses sources. La TSC-NHD est un projet multicentrique issu d'une collaboration de 22 centres médicaux se trouvant en majeure partie répartis dans 14 états des États-Unis. Deux centres ont depuis peu rejoint la TSC-NHD au Québec, le Centre hospitalier de l'Université de Montréal et le Centre hospitalier universitaire Sainte-Justine, bien que la réception des données utilisées dans ce projet de recherche précède leur inclusion. Des 22 centres participant à la base de données, 20 sont directement affiliés à une université et 8 sont des centres pédiatriques. Les objectifs poursuivis par la création de la base de données sont de documenter l'impact que peut avoir la maladie sur la santé d'une personne au cours de sa vie et de promouvoir la recherche portant sur la sclérose tubéreuse. Des individus consentants de tous âges avec un diagnostic définitif de STB actuellement suivis dans l'une des cliniques inscrites sont admissibles à rejoindre la TSC-NHD. Le recrutement des participants se fait selon une méthode d'échantillonnage non probabiliste de commodité. C'est-à-dire que les participants sont recrutés d'un centre à un autre sur une base volontaire, avec comme seule condition d'inclusion d'avoir reçu un diagnostic de STB.

L'information comprise dans la base de données a été recueillie selon une combinaison de méthodes prospectives et rétrospectives. Tous les renseignements pertinents précédant le recrutement d'un participant tel qu'un diagnostic obtenu durant l'enfance pour un adulte ont été ajoutés aux données de façon rétrospective au moment de l'inclusion. Suivant cette première entrée de données, une mise à jour avait lieu prospectivement lors des visites médicales subséquentes. Les données recueillies sont inscrites dans la base par un assistant de recherche employé dans chaque centre à partir de l'information contenue dans le dossier médical des participants. Celles-ci concernent les aspects sociodémographiques, les manifestations cliniques de la maladie comprenant l'âge au diagnostic et le niveau de sévérité, les tests de dépistage effectués ainsi que les traitements reçus.

Le projet a exploité les données allant de la création de la TSC-NHD en 2006 jusqu'à la date à laquelle elles nous ont été transmises, soit en juin 2022. À ce moment, de l'information sur 2420 individus avait été recueillie, ce qui en fait la plus grande base de données visant le suivi de personnes atteintes de la sclérose tubéreuse observée à ce jour dans la littérature.

## **4.3. Variables**

### **4.3.1. Prédicteurs**

Un total de douze prédicteurs potentiels a été considéré pour le développement des modèles. Ceux-ci ont été sélectionnés sur des bases de connaissances cliniques établies afin d'identifier les principales manifestations cliniques de la sclérose tubéreuse survenant durant l'enfance. Deux variables qu'on sait corrélées de près aux issues étudiées ont aussi été ajoutées, soit le sexe biologique et la mutation génétique en cause de la maladie. La liste complète des

prédicteurs accompagnée de leur description, des risques principaux auxquels on les associe et de leur opérationnalisation se trouve au Tableau 2.

L'information concernant le diagnostic des manifestations cliniques était entrée dans la base de données par un investigateur répondant à la question : « Est-ce que le participant a déjà souffert d'une des conditions suivantes préalablement à son inclusion dans la TSC Alliance ou depuis sa dernière visite ? ». Une variable dichotomique était ensuite créée, dénotant la présence ou l'absence de chacune des conditions. Le sexe biologique ainsi que la mutation génétique en cause de la maladie ont tous les deux aussi été traités en tant que variables dichotomiques. La première identifiant la personne comme étant un homme ou une femme et la seconde si le génotypage a permis de détecter une mutation du gène *TSC1* ou *TSC2*. La seule variable n'étant pas présentée sous une forme binaire est le degré de déficience intellectuelle. En effet, celui-ci peut prendre trois valeurs distinctes dénotant l'absence de déficience intellectuelle, un niveau léger à modéré de déficience intellectuelle ou un niveau de sévère à profond de déficience intellectuelle.

Tableau 2. – Prédicteurs considérés

<b>Variable</b>	<b>Description</b>	<b>Risques principaux</b>	<b>Opérationnalisation</b>
Sexe biologique	-	-	Hommes ou Femmes
Mutation génétique	Mutation en cause de la maladie ; <i>TSC1</i> , <i>TSC2</i> ou non identifiée	Risques plus élevés associés à la mutation <i>TSC2</i>	<i>TSC1</i> ou <i>TSC2</i>
Niveau de déficience intellectuelle	Présence de différents niveaux d'atteinte intellectuelle	Risques variant selon le niveau de sévérité de l'atteinte intellectuelle	Sévère ou Modéré ou Normal
Épilepsie	Présence de crises épileptiques	Risque élevé d'épilepsie réfractaire aux médicaments	Présence ou Absence
Dysplasies corticales	Malformation présente au cerveau	Augmentation du risque d'épilepsie	Présence ou Absence



Tumeurs sous-épendymaires à cellules géantes (SEGA)	Tumeur sous-épendymaire	Risque d'hydrocéphalie obstructive	Présence ou Absence
Nodules sous-épendymaires	Lésions sur les ventricules latéraux du cerveau	Risque de se développer en tumeurs sous-épendymaires à cellules géantes	Présence ou Absence
Angiofibromes faciaux	Lésions au visage s'apparentant à de l'acné	Pas de risque pour la santé physique. Risque de stigmatisation	Présence ou Absence
Macules hypopigmentées	Taches pâles pouvant apparaître partout sur le corps	Ne cause généralement aucun problème	$\geq 3$ ou $< 3$
Peau de chagrin	Surface de peau surélevée, généralement retrouvée sur le bas du dos	Ne cause généralement aucun problème	Présence ou Absence
Rhabdomyome cardiaque	Tumeur bénigne présente sur le coeur	Faible risque d'insuffisance cardiaque et d'arythmie	Présence ou Absence
Hamartomes rétiens	Tumeur bénigne de la rétine	Faible risque de perte de vision	Présence ou Absence

Tel que mentionné précédemment, l'aspect de la confusion n'est habituellement pas abordé lors du développement de modèles de prédiction étant donné que l'association entre les variables n'y est pas évalué. À cet effet, l'outil principal fréquemment utilisé en épidémiologie afin d'obtenir une visualisation graphique du lien causal unissant les variables, le graphique acyclique dirigé (DAG), n'est lui aussi que rarement aperçu dans ce type d'études [40]. C'est pourquoi nous avons pris la décision de ne pas inclure ce type de graphique dans le présent projet de recherche.

## 4.3.2. Issues d'intérêt

### 4.3.2.1. Angiomyolipomes rénaux

La première issue d'intérêt analysée était la présence d'un diagnostic d'angiomyolipomes rénaux. Les tumeurs retrouvées sur les deux reins et celles comprenant des lésions d'un diamètre de plus de trois centimètres sont fréquemment associées à un risque plus élevé de complications sévères, telles qu'un risque accru d'hémorragie intrapéritonéale ou rétropéritonéale [41]. Puisque les données de la TSC-NHD permettaient de faire cette distinction, deux issues supplémentaires concernant les AML ont été considérées. La première portait sur le caractère bilatéral ou unilatéral de l'AML et la seconde sur la présence d'AML contenant des lésions qui ont atteint une taille d'au moins trois centimètres. Tel qu'observé dans le Tableau 3, cela porte au nombre de trois les issues considérées en lien avec les AML.

Tableau 3. – Issues d'intérêt

Variable	Opérationnalisation
Angiomyolipomes rénaux	Présence ou absence d'au moins un angiomyolipome rénal
Angiomyolipomes rénaux bilatéraux	Présence ou absence d'angiomyolipomes sur les deux reins
Angiomyolipomes rénaux contenant des lésions d'au moins trois centimètres	Présence ou absence d'au moins un angiomyolipome rénal d'une taille d'au moins trois centimètres
Lymphangioléiomyomatose pulmonaire	Présence ou Absence de lymphangioléiomyomatose pulmonaire

L'information au sujet du diagnostic principal d'AML a été entrée dans la base de données par un investigateur suivant la question : « Est-ce que le participant a été diagnostiqué avec l'une des conditions suivantes ? ». La réponse obtenue a permis la création d'une variable dichotomique servant à déterminer si le participant avait reçu un diagnostic d'AML. Pour les deux issues

supplémentaires, l'investigateur devait répondre à des sous-questions spécifiant si l'AML diagnostiqué était bilatéral ou unilatéral, et si au moins une lésion d'une taille supérieure à trois centimètres avait été observée. Deux variables dichotomiques additionnelles ont été construites sur la base de ces informations. À noter que dans le cas des deux issues supplémentaires ajoutées, les participants qui ont un AML qui est présent seulement sur un rein ou ne contient que des lésions de moins de trois centimètres ont été considérés comme n'ayant pas développé l'issue.

#### **4.3.2.2. Lymphangioléiomyomatose pulmonaire**

La dernière issue d'intérêt est la présence de lymphangioléiomyomatose pulmonaire, qui comporte comme principaux risques pour la santé une dégradation de la fonction pulmonaire et le développement de pneumothorax [7]. Encore une fois, la variable concernant la détection de LAM a été créée d'après les réponses de l'investigateur à la question : « Est-ce que le participant a déjà eu une LAM ou une LAM probable précédant son inclusion dans la TS Alliance ou depuis sa dernière visite ? ». La variable résultante était présentée sous la forme dichotomique, indiquant la présence ou bien l'absence d'un diagnostic de LAM chez un participant.

#### **4.3.3. Critères d'exclusion**

Le choix des variables sélectionnées en tant que prédicteurs dans les modèles repose sur des bases de connaissances cliniques et sur l'étude de la littérature concernant la STB afin de ne sélectionner que des manifestations cliniques de la STB détectées chez des individus âgés de moins de 5 ans [7]. Un survol des données de la TSC-NHD a d'ailleurs pu confirmer que les manifestations cliniques s'étaient développées pour tous les participants avant l'âge de 5 ans. Il a été montré que les LAM affectent de façon quasi exclusive des femmes adultes, ce qui n'est toutefois pas le cas pour les AML. Effectivement, ceux-ci font leur apparition dans près de la moitié des cas durant l'enfance [7]. Afin d'assurer une temporalité adéquate entre les prédicteurs

et les issues, les analyses concernant les AML se sont limitées aux participants n'ayant toujours pas reçu de diagnostic une fois rendus à l'âge de 5 ans.

## **4.5. Analyses statistiques**

Toutes les analyses ont été effectuées en utilisant R V4.2 avec l'aide des bibliothèques suivantes : mice [42], miceadds [43], glmnet [44] et regression modeling strategies (rms) [45].

### **4.5.1. Préparation des données**

Plusieurs étapes de préparation des données ont été requises avant de pouvoir procéder aux analyses et au développement des modèles.

Premièrement, les données comportaient une grande quantité de variables qui ont été jugées comme n'étant pas utiles au présent projet de recherche. On y retrouvait entre autres des détails concernant l'ethnicité des participants, des renseignements supplémentaires au sujet des mutations génétiques, de l'information sur de nombreux troubles neurologiques dont souffrent les gens atteints de STB ainsi que des précisions à propos de la sévérité de certaines manifestations cliniques. Ces variables ont été retirées du jeu de données afin de ne conserver que celles utiles aux fins de cette étude.

La majorité des variables à inclure dans les analyses comprenaient une certaine quantité de valeurs manquantes. À cet effet, de multiples méthodes ont été développées afin de traiter les données manquantes avant de pouvoir procéder aux analyses. Parmi celles-ci, on retrouve l'analyse complète, qui consiste à ne retenir que les participants pour qui aucune donnée n'est manquante pour toutes les variables à inclure dans les modèles statistiques [46]. Cette technique, qu'on observe non seulement dans le développement de modèles de prédiction, mais aussi dans nombre d'autres études en épidémiologie, comporte cependant de nombreux désavantages. Elle

est en particulier dommageable si la taille de l'échantillon est petite et si la proportion de valeurs manquantes est élevée. En plus de la perte de précision et de puissance statistique associée à la réduction de la taille d'échantillon disponible, cette méthode tend d'ailleurs à mener à des estimations biaisées [46]. Un biais de sélection peut être causé par l'analyse des cas complets si les participants qui demeurent après le retrait ne sont pas représentatifs de la population étudiée. La direction ainsi que l'étendue du biais engendré sont difficilement prévisibles, puisqu'elles dépendent des raisons pour lesquelles les valeurs sont manquantes et de leur lien avec les issues étudiées [47, 48]. Une autre méthode visant à contrer les inconvénients résultant de l'analyse complète est l'imputation multiple par équations chaînées. Celle-ci consiste à créer un nombre prédéfini d'échantillons plausibles ne contenant aucune donnée manquante. Le processus d'imputation multiple comprend plusieurs étapes [49] :

1. Chaque valeur manquante est temporairement remplacée à l'aide d'une technique d'imputation simple, généralement le remplacement par la moyenne.
2. Les valeurs temporaires sont retirées pour l'une des variables.
3. Cette variable qui contient à nouveau des valeurs manquantes est utilisée en tant que variable dépendante d'un modèle de régression qui prend pour variables indépendantes toutes les autres ne contenant maintenant plus de valeurs manquantes.
4. Les résultats du modèle de régression sont utilisés afin de prédire les valeurs pour la variable pour laquelle le remplacement temporaire avait été retiré.
5. Les étapes 2 à 4 sont répétées pour chacune des variables restantes comprenant des valeurs manquantes, en utilisant les variables imputées par la régression plutôt que celles par le remplacement par la moyenne dès qu'elles sont disponibles.

6. Les étapes 1 à 5 mises ensemble constituent un cycle ou une itération. Elles sont ensuite répétées pour un nombre de cycles prédéfini, en utilisant à l'étape 1 les valeurs obtenues à la fin du cycle précédent plutôt que celles provenant de l'imputation par remplacement par la moyenne.
7. L'entièreté du processus est ensuite répétée pour chaque jeu de données imputé que l'on souhaite obtenir.

Par la suite, les résultats des analyses obtenus de ces multiples échantillons sont mis en commun afin d'être interprétés [50]. Lorsqu'utilisée sur des données pour lesquelles on suppose que les valeurs manquantes le sont de manière soit complètement aléatoire ou simplement aléatoire, cette méthode tend à produire des estimations non biaisées tout en permettant de conserver chacun des participants inclus dans l'échantillon initial [49].

Nous avons cru raisonnable d'émettre l'hypothèse que les valeurs manquantes dans la TSC-NHD le sont de manière aléatoire. C'est-à-dire que bien que les observations manquantes ne le soient peut-être pas de façon entièrement aléatoire, leur absence peut tout de même être pleinement expliquée par l'information présente dans les données [50]. À la lumière de cette information et en prenant en considération les divers avantages qu'elle confère, nous avons opté pour l'imputation multiple par équations chaînées afin de traiter les valeurs manquantes dans nos données. Un total de dix échantillons complets a été créé avec un nombre maximum d'itérations fixé à 20. Tous les prédicteurs ainsi que les issues ont été intégrés dans le processus d'imputation, comme indiqué dans les plus récentes recommandations [50, 51]. La régression polytomique a été choisie pour la variable concernant le degré de déficience intellectuelle étant donné ses trois niveaux possibles, tandis que la régression logistique fut sélectionnée pour les autres variables, qui sont toutes de forme dichotomique.

## **4.5.2. Description des caractéristiques des participants**

Tout d'abord, une analyse descriptive des données a été effectuée. Les résultats de cette analyse ont pu être résumés dans un tableau divisé en deux parties, l'une pour les analyses concernant l'AML et l'autre pour celle concernant la LAM. Cette distinction découle des tailles d'échantillons qui diffèrent entre les analyses portant sur les AML et celles sur les LAM. Le nombre ainsi que la fréquence en pourcentage respective pour l'ensemble des prédicteurs et des issues ont été rapportés. La quantité d'observations manquante pour chaque variable  $y$  est aussi présentée, avec encore une fois le pourcentage associé.

## **4.5.3. Analyses principales**

Tout d'abord, étant donné que quatre issues d'intérêt ont été considérées dans ce projet de recherche (trois pour AML et une pour LAM), les analyses concernant chacune d'entre elles ont été menées séparément. À cet effet, chacune des étapes du développement des modèles, de la sélection des variables, de la construction des scores de risque clinique, de la validation et de l'évaluation de la performance ont été appliquées tour à tour en suivant les mêmes procédures présentées dans cette section pour chacune des issues.

La régression de Cox ainsi que la régression logistique sont toutes les deux utilisées dans le développement de modèles de prédiction [52]. La régression de Cox a pour avantage de tenir compte de la durée de survie avant qu'une issue se produise ainsi que de considérer la censure chez les participants. Ces deux aspects permettent de retirer plus d'information des données que ce qui est habituellement possible avec la régression logistique. Cependant, les données de la TSC-NHD ne contenaient pas de renseignements concernant la date de diagnostic pour les AML bilatéraux ainsi que pour les AML avec lésions de plus de trois centimètres. De plus, une date de

fin de suivi se doit d'être présente pour chaque participant et correspond généralement à la survenue de l'issue, à la fin de l'étude ou bien à la censure d'un participant. Cette variable garantit qu'au moins jusqu'à cet instant, l'événement étudié ne s'était pas produit pour ceux n'ayant pas développé l'issue. Toutefois, dans la TSC-NHD, la seule variable pouvant servir à cet effet correspondait à la date à laquelle la plus récente mise à jour avait été faite pour un participant dans les données. Rien ne nous indique cependant qu'un suivi a été effectué à ce moment afin de déterminer par exemple si un participant avait développé un AML depuis son dernier rendez-vous médical. Nous n'aurions donc pas été en mesure de garantir l'absence de l'issue selon les variables offertes par la TSC-NHD, menant à une durée de suivi biaisée qui aurait pu nuire aux estimations faites par les modèles de prédiction. À ces aspects s'ajoute un grand nombre de valeurs manquantes pour les dates de diagnostic d'AML et de LAM, près de la moitié des diagnostics n'étant pas accompagnés d'une date. Ce problème aurait potentiellement pu être réglé par l'imputation multiple, mais étant donné la qualité générale des variables concernant les dates, le risque de biais demeurait relativement élevé.

Pour les raisons énumérées dans le paragraphe précédent, nous avons opté pour la régression logistique dans l'élaboration des modèles pour les quatre issues étudiées. L'utilisation de ce type de régression nous a permis d'observer l'apparition des issues pendant le suivi des participants, sans information sur la durée de suivi la précédant. Dans un premier temps, un modèle complet comportant chacun des douze prédicteurs potentiels a été développé à partir des résultats combinés provenant des dix jeux de données imputés. Ceux-ci ont permis d'obtenir dans chaque cas une estimation des coefficients de régression et du rapport de cotes (OR) représentant le risque de développer l'une des issues selon la présence ou l'absence des douze prédicteurs compris dans



les modèles. Les intervalles de confiance à 95 % ont aussi été calculés pour les OR afin d'évaluer la précision des estimations.

Nous avons pris la décision d'inclure un second modèle dans lequel un processus de sélection des variables a été déployé. Appliqué à chacune des quatre issues, cela porte le nombre total de modèles qui ont été conçus dans le cadre des analyses principales à huit, soit deux pour chacune d'entre elles. L'objectif premier de l'emploi d'une méthode de sélection des variables était d'obtenir un second modèle contenant moins de prédicteurs, permettant ainsi de simplifier son utilisation future dans un contexte clinique. L'usage d'une méthode de sélection des variables se produisant durant le processus de développement peut mener à un modèle adapté de trop près aux données, ayant pour conséquence une diminution de la performance prédictive [23, 24]. Une attention particulière a été portée à ces aspects, tant dans le choix de la méthode de sélection que dans son application, ainsi que dans l'évaluation subséquente de la performance des modèles.

Une approche de sélection des prédicteurs fréquemment utilisée est la sélection par étapes descendante ou ascendante (*backward and forward stepwise*). La sélection par étape descendante consiste à déterminer de prime abord un seuil d'arrêt appuyé dans la plupart des cas sur une valeur- $p$  ou sur l'AIC (*akaike selection criteria*). En débutant avec le modèle complet, les variables ne respectant pas ce seuil préétabli sont retirées une à une. Le tout est répété jusqu'à ce que tous les prédicteurs qui demeurent correspondent au seuil choisi [53]. L'utilisation de ce type de procédé automatisé reposant sur la signification statistique tend néanmoins à produire une estimation des coefficients de régression biaisée et à offrir une évaluation de la performance trop optimiste. Il peut aussi conduire au rejet de prédicteurs contenant de l'information pertinente aux analyses étant menées [54]. Une autre option permettant de procéder à la sélection des variables est le LASSO (*least absolute shrinkage and selection operator*) qui, comme son nom l'indique, est une méthode

de contraction des coefficients [31]. Le LASSO consiste à ajouter au modèle de régression un terme de pénalité accompagné d'un paramètre de régularisation représenté par le symbole  $\lambda$ . Ce paramètre choisi par l'utilisateur correspond au niveau de pénalisation qui est appliqué aux coefficients de la régression. L'insertion de ce terme de pénalisation dans le modèle de régression complet réduit la valeur des coefficients du modèle, certains pouvant aller jusqu'à atteindre une valeur de zéro. Seules les variables pour lesquelles le coefficient est demeuré supérieur en valeur absolue à zéro sont conservées dans le modèle de régression LASSO. Parmi les avantages principaux de cette méthode de sélection, la pénalisation qui est appliquée aux coefficients contribue à limiter le risque de produire des estimations trop optimistes. La performance des modèles LASSO telle qu'évaluée à l'aide de mesures de calibration et de discrimination tend aussi à être semblable à celle d'un modèle complet [30].

Nous avons donc opté pour la sélection des variables avec le LASSO pour chacune des quatre issues. Pour ce faire, le modèle complet incluant les douze prédicteurs a été évalué dans les dix bases de données imputées individuellement. Ensuite, un processus de validation croisée a été appliqué afin d'extraire pour chaque base de données une valeur du paramètre de régularisation optimale. La moyenne des dix  $\lambda$  obtenus a été calculée puis ajoutée à un modèle complet développé dans une base de données superposant les dix imputations. Les coefficients ayant atteint une valeur de zéro ont été retirés de leurs modèles respectifs et ceux qui demeuraient ont été présentés avec les OR associés.

#### **4.5.4. Validation et évaluation de la performance**

Une validation interne a été effectuée pour chacun des huit modèles développés à l'aide desquels des mesures de performance corrigées pour l'optimisme ont pu être extraites. La validation par partition d'échantillon (*split-sample*), la validation croisée et le rééchantillonnage

par bootstrap figurent parmi les trois approches fréquemment utilisées pour valider un modèle de prédiction sans avoir accès à des données externes [28, 32].

La partition d'échantillon consiste à diviser préalablement au développement des modèles l'échantillon en deux parties de façon aléatoire. Le modèle est ensuite développé dans la première partie et sa performance est évaluée dans la deuxième [32]. Cette approche a pour avantages d'être simple à employer et rapide à appliquer. En revanche, en plus des pertes de puissance statistique associées à la séparation de l'échantillon en deux sections, cette technique tend à produire des estimations de la performance instables comportant un haut niveau de variabilité [28].

En ce qui a trait à la validation croisée, celle-ci peut s'opérationnaliser selon plusieurs variations distinctes. L'élément principal reliant chacune d'entre elles est la répétition sur plusieurs itérations d'une méthode demandant l'utilisation de différentes parties de l'échantillon pour développer le modèle et pour en évaluer sa performance. Une des approches les plus populaires est la validation croisée à  $k$ -blocs (*k-fold cross validation*). Elle consiste à diviser l'échantillon total en un nombre  $k$  de blocs, à développer le modèle dans l'un d'eux et à en évaluer la performance dans les  $k-1$  restants. Le processus est par après réitéré jusqu'à ce que chacun des blocs ait servi au développement du modèle, menant à l'obtention d'une quantité de mesures de performance équivalente au nombre de blocs sélectionnés. La moyenne ainsi que l'écart-type des mesures de performance peuvent ensuite être calculés. Cette approche est simple à appliquer et permet l'utilisation de l'entièreté de l'échantillon initial afin d'obtenir une estimation des mesures de performance avec un faible niveau de biais et de variabilité [28].

La validation d'un modèle et l'évaluation de sa performance avec la méthode de rééchantillonnage par bootstrap comportent plusieurs étapes. Dans un premier temps, le modèle de prédiction est développé dans la base de données d'origine et sa performance apparente est

mesurée. Elle est dite apparente puisqu'elle correspond à une performance évaluée dans les mêmes données ayant servi au développement d'un modèle. Elle n'est par conséquent pas représentative de sa performance réelle lorsque confrontée à d'autres données. La deuxième étape consiste à créer un échantillon de bootstrap en effectuant une pige aléatoire avec remise dans la base de données jusqu'à l'obtention d'un échantillon de taille égale à cette dernière. Un modèle de prédiction est ensuite développé dans ce nouvel échantillon en prenant soin de répéter toutes les étapes du processus de sélection des variables qui ont été employées dans l'élaboration du modèle original. La performance de ce modèle est appelée la mesure de performance bootstrap. L'étape suivante consiste à appliquer le modèle conçu dans l'échantillon bootstrap aux données d'origines et d'évaluer encore une fois la performance qu'on nomme la performance de test. Le niveau d'optimisme, qui correspond à la surévaluation des capacités prédictives d'un modèle, peut par la suite être quantifié en soustrayant la performance de test à la performance bootstrap. Toutes les étapes à partir de la création de l'échantillon bootstrap sont répétées un nombre  $n$  de fois, ce qui permet d'assurer la stabilité de la mesure d'optimisme. Bien qu'il n'existe pas de recommandation unique quant à la quantité de répétitions à appliquer, un minimum de 100 est en règle générale observé. Pour terminer, la moyenne des  $n$  mesures d'optimisme est calculée et est soustraite à la performance apparente, ce qui mène à l'obtention d'une mesure validée de la performance corrigée pour l'optimisme [24, 30, 55]. Cette approche comporte des avantages semblables à ceux de la validation croisée, présentant un niveau de biais et une variabilité vraisemblablement équivalentes [28].

Différentes mesures de la performance peuvent être extraites lors du processus d'évaluation de la performance prédictive d'un modèle. La capacité discriminative d'un modèle de prédiction est la plupart du temps estimée à l'aide du c-index. Celui-ci permet de déterminer dans quelle

proportion le modèle développé est en mesure de distinguer entre une personne qui développera l'issue et une autre qui ne la développera pas [32]. Le c-index peut en théorie prendre des valeurs allant de 0 à 1, bien qu'en pratique il se situe plus généralement entre 0,60 et 0,85 [23]. Une valeur de 0,5 pour le c-index représente une capacité discriminative d'un modèle qui est équivalente au hasard tandis qu'une valeur de 1 indiquerait plutôt que le modèle arrive à discriminer de façon parfaite entre les individus qui développeront l'issue et ceux qui ne la développeront pas. Pour ce qui est de la calibration, on l'illustre en règle générale à l'aide d'une courbe dans laquelle les probabilités de survenue de l'issue prédites par le modèle sont comparées à celles réellement observées dans les données [23]. Une dernière mesure qui accompagne souvent la discrimination et la calibration d'un modèle est le  $R^2$  de Nagelkerke. Celui-ci permet de déterminer la proportion de la variation totale du risque de développer l'issue pouvant être expliquée par les prédicteurs inclus dans le modèle [23].

Nous avons opté pour le rééchantillonnage par bootstrap dans la validation de nos modèles de prédiction afin d'obtenir des mesures de performance fiables et corrigées pour l'optimisme. Le processus de validation interne a été effectué dans chacune des dix bases de données imputées dans le cas des modèles complets comprenant tous les prédicteurs. Dans chaque cas, le modèle a été validé sur un total de 300 répétitions. Cette méthode nous a permis d'extraire les principales mesures de performance, évaluées indépendamment dans chaque base de données imputée, que nous avons jugées pertinentes à l'estimation de la capacité prédictive des modèles, soit le c-index, les courbes de calibration et le  $R^2$  de Nagelkerke. Les moyennes des valeurs du c-index et du  $R^2$  de Nagelkerke sur les dix bases de données ont été calculées et sont présentées dans les résultats afin d'avoir une meilleure perception de la performance globale des modèles. Pour ce qui est des

courbes de calibration, elles sont fournies pour chaque modèle selon la moyenne provenant des dix bases de données imputées.

Pour les modèles dans lesquels une sélection des prédicteurs par la méthode du LASSO a été employée, toutes les étapes de sélection des variables telles qu'elles sont décrites plus haut ont été reproduites afin de valider les modèles avec un total de 100 répétitions du rééchantillonnage par bootstrap. Un nombre de répétitions inférieur à celui retrouvé dans le modèle complet a été choisi en raison de la demande informatique considérable associée à la répétition de toutes les étapes de sélection LASSO. Le processus de validation interne a encore une fois été appliqué à chacune des dix bases de données imputées. Tout comme pour les modèles complets, des mesures du c-index et du  $R^2$  ont pu être extraites et les moyennes des valeurs provenant des dix bases de données ont été présentées dans les résultats.

#### **4.5.5. Présentation des résultats**

Les résultats complets des régressions incluant la valeur de tous les coefficients avec l'ordonnée à l'origine ont été fournis, selon les recommandations du *Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis* (TRIPOD) [20]. Pour la présentation finale, nous avons opté pour l'élaboration d'un score de risque clinique. Cette approche de présentation des résultats a pour avantage d'être relativement simple à comprendre et à utiliser dans un contexte clinique une fois développée. De plus, puisque nos modèles ne contenaient que des variables catégorielles et qu'ils ont été développés à l'aide de régressions logistiques, ils sont particulièrement bien adaptés au développement de scores de risque clinique. En effet, les principaux inconvénients de ce type de présentation des résultats sont l'obligation de catégoriser les prédicteurs continus et la sélection d'une durée de suivi d'intérêt lorsqu'ils sont appliqués à des issues de survie [33].

Le score de risque clinique pour chacun des huit modèles a été présenté sous la forme de deux tableaux et d'un graphique. Le premier tableau dépeint le score rattaché à chaque prédicteur et le second la probabilité en pourcentage de développer l'issue selon le score total mesuré pour un individu. Le graphique quant à lui, permet une visualisation rapide des résultats présentés dans le second tableau. La première étape de l'élaboration du score consiste à attribuer un poids à chaque prédicteur [24, 33]. Le calcul des poids s'est fait en divisant les coefficients non transformés de la régression logistique associés à chaque prédicteur par la plus petite valeur absolue de tous les coefficients du modèle. Le tout a par la suite été arrondi à la première décimale. Le score total pour un individu équivaut donc à la somme des poids rattachés aux conditions médicales diagnostiquées, à laquelle s'ajoute le score relié à son sexe biologique, à son niveau de déficience intellectuelle et à la mutation génétique en cause de la maladie. La seconde étape consiste à déterminer, pour chaque valeur du score, la probabilité en pourcentage de développer l'issue. Pour ce faire, le score total pour chaque participant dans la base de données a été calculé en multipliant la matrice des valeurs pour chaque prédicteur (illustrés de façon dichotomique avec une valeur de 1 dénotant la présence et une valeur de 0 indiquant l'absence du prédicteur) par le vecteur des scores associés à chaque prédicteur. Ensuite, le score total par participant est placé en tant que variable indépendante dans un modèle de régression logistique ayant pour variable dépendante l'issue du modèle original. Le coefficient résultant de cette régression représente la valeur qui est multipliée au score total pour établir le risque. L'ordonnée à l'origine du modèle de score est déterminée en insérant dans un modèle de régression logistique avec l'issue d'origine du modèle complet un « *offset* » correspondant à la multiplication du coefficient identifié à l'étape précédente par le score total par individu. La valeur *log-odds* de survenue de l'issue est ensuite transposée en probabilité à l'aide de la fonction :

$$\text{Probabilité de survenue de l'issue} = \frac{\exp(\ln(odds))}{1 + \exp(\ln(odds))}$$

En résumé, un second modèle de régression logistique dans lequel la variable indépendante est le score total pour chaque participant et la variable dépendante est l'issue d'origine étudiée a été développé. La probabilité de survenue de l'issue a par la suite été calculée à partir de celui-ci. Un tableau a par après été construit afin de présenter l'étendue des probabilités de survenue de l'issue selon le score total.

Le c-index apparent a été mesuré pour chaque modèle de score de risque clinique développé. Cette mesure a été comparée à celle du modèle de régression initial afin de quantifier la perte de performance prédictive résultant de l'arrondissement des coefficients de régression.

#### **4.5.6. Analyses de sensibilité**

Des analyses de sensibilité ont été menées pour chacune des quatre issues étudiées afin de déterminer l'impact de l'âge d'un participant sur les estimations des modèles et sur leur performance prédictive. Un total de trois modèles supplémentaires a été développé pour chaque issue. Le premier limitait l'investigation aux participants ayant atteint l'âge de 12 ans lors de leur plus récente visite de suivi, le second aux participants ayant atteint l'âge de 18 ans et le dernier comprenait seulement les participants de 30 ans et plus. Les modèles ont été conçus au moyen de la régression logistique et incluaient les mêmes variables que les modèles complets dans l'analyse principale. Les c-index et les  $R^2$  de Nagelkerke apparents et corrigés pour l'optimisme à l'aide de la validation par bootstrap avec 300 répétitions ont été évalués pour l'entièreté des modèles (le processus est décrit en plus grand détail à la section 4.5.4.). Les résultats de ces modèles ont été présentés sous la forme d'un tableau par issue en illustrant les OR avec leur intervalle de confiance à 95 % ainsi que les mesures de performance prédictive.



## **4.6. Considérations éthiques**

Le présent projet de recherche a reçu l'approbation du comité d'éthique de la recherche du Centre hospitalier de l'Université de Montréal. Tous les participants recrutés dans la TSC-NHD ont rempli un formulaire assurant leur consentement libre et éclairé préalablement à leur inclusion dans l'étude. Toutes les données comprises dans le TSC-NHD sont dénominalisées, rendant impossible l'identification des participants dans la base de données. De plus, seuls les membres de l'équipe de recherche ont eu accès aux données, et ce uniquement à des fins de recherche. Seuls les résultats finaux des analyses seront publiés dans un journal révisé par les pairs.

## Chapitre 5 — Résultats

La section des résultats du projet de recherche sera présentée sous la forme d'un manuscrit d'article qui sera soumis prochainement pour publication au *Journal of Pediatrics*.

### 5.1. Article

**TITLE:** Development and internal validation of a clinical risk score to predict incident renal and pulmonary tumours in people with tuberous sclerosis complex.

#### Authors

Frédéric Loubert BSc,<sup>1,2</sup> Andrew A. House MD,<sup>3</sup> Catherine Larochelle PhD<sup>1,4</sup>, Philippe Major MD,<sup>4,5</sup> Mark R. Keezer PhD<sup>1,2,4</sup>

#### Affiliations:

1. CHUM Research Centre, Université de Montréal, Montreal, Canada.
2. School of Public Health of the Université de Montréal, Department of Social and Preventive Medicine, Montreal, Canada.
3. Division of Nephrology, Schulich School of Medicine & Dentistry, Western University and London Health Science Centre, London, Canada
4. Department of Neurosciences, Université de Montréal, Canada
5. Research Centre of the Centre Hospitalier Universitaire de Sainte Justine, Montreal, Canada.

#### Corresponding author:

**Mark R Keezer**

Centre Hospitalier de l'Université de Montréal (CHUM)

1000 rue Saint Denis

Montreal, QC Canada, H2X 0C1

+1 (514) 890-8233

mark.keezer@umontreal.ca

**Funding/Support:** TD bank Ready Commitment program

**Data sharing:** Data may be obtained from the TSC Alliance through the research data request form (<https://www.tscalliance.org/researchers/natural-history-database/>).

**Disclaimer statement:** The views expressed in this article are those of the authors and do not necessarily reflect the opinion of the TSC Alliance.

**Short title:** Prediction of tumours in people with TSC.

**Keywords:** TSC, angiomyolipoma, lymphangioliomyomatosis, prediction

**Abbreviations:** AML: Angiomyolipoma

LAM: Lymphangioliomyomatosis

LASSO: Least absolute shrinkage and selection operator

MRI: Magnetic resonance imaging

OR: Odds ratio

TRIPOD: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis

TSC: Tuberous sclerosis complex

TSC-NHD: TSC Natural History Database

## **ABSTRACT**

### **Objective**

Develop and internally validate a clinical risk score to predict incident renal AML and pulmonary LAM in people with tuberous sclerosis complex.

### **Study design**

Data from 2,420 participants in the TSC Alliance Natural History Database were leveraged for these analyses. Logistic regression was used to predict AML and LAM development using 10 clinical manifestations of TSC as potential predictors, in addition to sex and genetic mutation. For our models, we divided AML into three separate outcomes; presence or absence of AML, unilateral or bilateral, as well as whether any are  $\geq 3$  cm in diameter. The resulting regression models were turned into clinical risk scores which were then internally validated using bootstrap resampling, measuring discrimination and calibration.

### **Results**

The lowest clinical risk scores predicted a risk of AML and LAM of 5% and 0%, while the highest scores predicted a risk of 96% and 54%, respectively. Calibration was excellent for all three AML outcomes and good for LAM. Discrimination ranged from good to strong. C-statistics of 0.79, 0.79, 0.75, and 0.85 were seen for AML, bilateral AML, AML with a lesion  $\geq 3$  cm, and LAM, respectively.

### **Conclusion**

Our work is an important step towards identifying individuals who would benefit from prophylaxis and towards leading to more tailored schedules for screening imaging. We expect that our work will allow for more personalized medicine in people with TSC. External validation of the risk scores will be important to confirm the robustness of our findings.

## **INTRODUCTION**

Tuberous sclerosis complex (TSC) is a genetic disease which affects approximately one in 6000 live births [1]. The primary sources of morbidity for children with TSC are manifestations affecting the heart, brain, and skin. Adolescents and adults, on the other hand, are at greater risk of developing renal angiomyolipomas (AML) and pulmonary lymphangiomyomatosis (LAM) [1].

There is dramatic phenotypic heterogeneity in people with TSC, which can be seen even among members of the same family presenting with the same genetic mutation [2]. The burden of illness is also influenced by the age of the individual [1]. Currently, little information is available to help clinicians determine which of their patients affected by TSC are at greater risk of developing AML or LAM. The published literature is limited to showing that increasing age, female biological sex, and mutations of the TSC2 gene (as compared to TSC1) are associated with an increased risk of AML and LAM [3-5]. Knowledge of these risk factors does not allow for imaging schedules (e.g., MRI) that are more adapted to a person's individual risk. As a result, screening recommendations treat every individual equally and ignore the significant heterogeneity of the disease, making them subject to lifelong serial imaging [8]. In addition to personalizing imaging schedules, the identification of high-risk individuals could prove useful in targeting those most likely to benefit from prophylactic treatments, including mTOR inhibitors. It is currently unknown whether additional and more discriminating clinical features occurring during childhood can help identify adolescents and adults of low versus high risk of developing AML or LAM.

This study aimed to develop and internally validate a clinical risk score to predict incident renal AML and pulmonary LAM in people with TSC.

## **METHODS**

We report our study according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement [20].

### **Data sources**

Our study used data from the TSC Natural History Database (TSC-NHD), an ongoing clinical database implemented by the TSC Alliance in 2006. The TSC-NHD is a collaboration of 22 medical centres from across 15 states in the USA and 1 Canadian province. Consenting individuals of all ages with a definite TSC diagnosis who are followed at one of the 22 clinical sites are eligible to participate.

The TSC-NHD collects information on participants' sociodemographics, genotype, clinical disease manifestations (including their date of diagnosis and severity), imaging results and previous treatments received. Data collection comes from a combination of prospective and retrospective methods. Once participants are recruited, information preceding their entry is entered into the database. Following the first data entry, information is then updated at each subsequent clinical visit.

Data from the creation of the TSC-NHD in 2006 up until when access to the data was acquired (June 2022) were used in our analyses.

### **Outcomes**

The first outcome of interest was the development of renal AML, a form of benign tumour found on the kidneys. They are usually asymptomatic, but can sometimes lead to life-threatening complications, such as renal artery aneurysms and spontaneous rupture with intraperitoneal hemorrhage [1]. Risks of complications are higher for tumours with lesions greater than three cm and when they are present on both kidneys [41]. To account for those increased risks of complications, renal AMLs were characterized as three separate outcomes. The first one compared the presence to the absence of AML, the second whether they affected one kidney (i.e. unilateral) or bilateral kidneys, and the last one if they contained lesions greater than three cm at their largest diameter or not.

The last outcome of interest was the development of pulmonary LAM, a form of tumour found on the lungs almost exclusively in women of childbearing age [1]. Complications of LAM range from shortness of breath to respiratory failure [7].

The four outcomes were identified in participants either before or during their follow-up in the TSC-NHD and were captured in the database by investigators answering the question: "Has the participant been diagnosed with any of the following conditions?". All four outcome variables are dichotomous, denoting either the specified condition's presence or absence.

### **Predictors**

Twelve predictors of interest were considered in the development of the prognostic models. All of them were selected due to their being the major clinical manifestations of TSC occurring in children [1, 7]. Major brain, cutaneous, and heart manifestations were included, along with the biological sex of the participants and the responsible genetic mutation (TSC1 versus TSC2). To ensure appropriate temporality between the outcomes and the predictors, only clinical

manifestations diagnosed before the age of 5 years were considered [7]. Renal angiomyolipomas (AML) and pulmonary lymphangiomyomatosis (LAM) develop in adolescence and adulthood, not childhood, in most individuals [7]. A detailed list of all the predictors considered, along with the four outcomes is included in Table 1.

Clinical manifestations were identified either before or during a participant's follow-up. They were captured by investigators answering the question: "Has the participant ever had any of the following conditions before enrolment in the TS Alliance database or since last visit?". For all of them, a dichotomous variable was created, denoting either the absence or the presence of each condition.

### **Statistical analysis**

Data on all participants included in the database were used in model development to maximize the generalisability of the results. Missing data were present for most variables, with degrees of missingness ranging from 4.64 to 18.05 % (Table 1). Multivariate imputation by chained equations was used to handle missing data by creating 10 complete datasets [42]. All candidate predictors as well as the outcomes were included in the multiple imputation process [46, 50, 51]. Regression estimates and standard errors were pooled, while performance measures were evaluated in every imputed dataset and then combined [56].

Four distinct prognostic models were created, one for each outcome (three AML models, and one LAM model), using multivariable logistic regression. All predictors were already dichotomized, apart from the variable relating to intellectual disability, which was transformed into dummy variables using normal intellectual ability as the reference category. For each outcome, a full model including the 12 candidate predictors was developed, along with a second model in which variables were selected using the LASSO procedure [31]. The procedure was applied using cross-validation over the ten imputed datasets to identify the optimal penalty term value. This term was applied to the full regression models, penalizing their coefficients, and forcing some of them to a value of zero. The corresponding predictors were then removed, resulting in simplified models.

A clinical risk score was developed for each model by assigning weights to each predictor [24, 33]. The weights were calculated by dividing the untransformed log-odds regression coefficients by the smallest absolute value of all coefficients, rounded to the first decimal. Since all predictors were categorical, the score for a participant is equal to the sum of the weights for all the medical conditions for which they were diagnosed.

Internal validation and evaluation of the predictive performance of all models were assessed using a bootstrapping procedure [28]. Contrary to other methods such as split-sample, bootstrapping allows the use of the full cohort for both the development and the validation of the models. This allows for a reduction of overfitting, more stable estimates, and a lower chance of bias [28]. Interval validation was performed using bootstrapping with 300 resamples for the full models and 100 resamples for the LASSO models, and in both cases all model development steps were repeated in the bootstrapping process [57]. Although a higher number of resamples would have been preferable for the LASSO, we limited ourselves to 100 due to how computer-intensive the procedure is.

Model prognostic performance was assessed using discrimination and calibration measures [32, 58]. Discrimination refers to the capacity of a model to differentiate between those who develop the outcome and those who do not and is operationalised using Harrel's C-statistic. A C-statistic value of 0.5 represents a capacity to discriminate that is not superior to chance, while a value of 1.0 represents perfect discrimination. Models with a C-statistic value above 0.7 are typically interpreted as good, while models with values above 0.8 are seen as having strong discriminative ability [59]. Calibration refers to the agreement between outcomes predicted by the models and observed outcomes in the data. It is often presented using a calibration curve, in which perfect calibration is indicated by a slope of 1. A model's performance evaluated using the same data used for its development is known as apparent performance, while performance evaluated using either external data or an internal validation procedure is known as optimism-corrected performance. Both apparent and optimism-corrected estimates of performance were reported for every model. Apparent C-statistics for the rounded scores were calculated to evaluate the loss of performance arising from the rounding of the regression coefficients.

We carried out sensitivity analyses in which all full models were re-evaluated based on participants' ages at their last follow-up visit. The first sensitivity analysis included only participants older than 12 years at their last visit, the second only participants older than 18, and the last only participants older than 30. This allowed an investigation into the potential impact of participant age on model performance.

All analyses were performed using R V.4.2., along with the R libraries mice [60], miceadds [43], glmnet [44], and rms [45].

### **Inclusion/Exclusion criteria**



Of the 2,420 participants included in the TSC-NHD, 284 were excluded from the AML analyses because they had received a diagnosis of AML before the age of 5 years. A further 197 participants were also removed because their last follow-up visit occurred before the age of 5. The remaining 1,939 participants were included in the development and validation of the AML models. As temporality between predictors and outcome was not an issue for the LAM models (all LAM were diagnosed after age 5), all 2,420 participants in the TSC-NHD were included in both the development and validation of the model.

## **RESULTS**

Participant sociodemographic characteristics and their clinical features, along with the distribution of missing data are shown in Table 1.

Using our clinical tools, an individual's risk is estimated by calculating the sum of scores associated with each TSC feature present, along with the sex and the responsible genetic mutation (Table 3). The probability of developing the outcome associated with an individual's risk score is obtained from Table 4 and Figure 2.

Regression coefficients, along with their respective ORs and 95% confidence intervals (95% CI) for the full models are shown in Table 2.

### **AML**

After imputation, 1,000 participants developed an AML during their follow-up. When converted to a clinical risk score, our regression model was able to predict risks of eventually developing AML from as low as 5% to risks as high as 96%. The overall performance of this model was good. The C-Index showed a good ability for the model to discriminate between individuals who have and those who do not have AML while the calibration plot showed a strong concordance between the probabilities of developing an AML seen in our sample and the probabilities predicted by our model. More detailed information about the model's performance is found in Table 2 and Figure 1.

When applying the LASSO procedure to the AML model, four predictors were removed from the model (moderate intellectual disability, cortical tuber hamartoma, epilepsy, and cardiac rhabdomyoma). The removal of predictors led to a smaller range of risk prediction (Online supplement).

### **BILATERAL AML**

This second outcome compared participants who developed a bilateral AML during follow-up with participants who either had not developed an AML or had a unilateral AML. After imputation, 792 participants developed bilateral AML during follow-up. Predictions for bilateral AML ranged from a 2% risk to a 93% risk of development. Model performance was comparable to what was seen for the first outcome, with good discrimination and great calibration.

The regression coefficients for the variables for moderate intellectual disability, cortical tuber hamartoma, epilepsy, and hypomelanotic macules all shrank to zero in the LASSO model, which led to their removal. This simplified model suffered a small loss in its range of risk prediction (Online supplement).

### **AML LESIONS GREATER THAN 3 CM**

The third outcome compared participants who developed an AML containing lesions with a diameter greater than 3 centimetres with participants who did not. Following imputation, 1939 participants were included with 378 of them developing the outcome. Using the available predictors, risks as low as 1% and as high as 72% of eventually developing an AML with lesions greater than three cm can be predicted. The predictive performance for this model is good, as seen in Table 2 and Figure 1.

A total of five variables shrank to zero when applying a LASSO procedure to the model (moderate intellectual disability, severe intellectual disability, subependymal nodule, cortical tuber hamartoma, and epilepsy). As seen in the other models, the LASSO led to a reduction in the range of predictions (Online supplement).

### **LAM**

One hundred and seven participants developed LAM during their follow-up. The range of predictions was considerably smaller than what was seen in the AML models, with a low of 1% and a high of 55%. The model performance showed a strong ability to discriminate. The calibration plot showed a modest level of concordance, which tended to diverge for predicted probabilities greater than 20%.

Eight variables were excluded from the LASSO model, due to their coefficients shrinking to zero, leaving only biological sex, genetic mutation, epilepsy, facial angiofibroma, and cardiac rhabdomyoma. The removal of those predictors had a considerable impact on the range of predictions, with the highest predictable risk going from a 55% chance to a 27% chance of eventually developing AML.

## **SENSITIVITY ANALYSES**

For the three AML models, limiting the analyses to individuals who were at least 12 years old at their last follow-up decreased the number of participants from 1939 to 1564, limiting to 18 years or more dropped the number of participants to 1153 and limiting to 30 years or older left 553 participants. For the LAM model, the number of participants went from 2420 to 1750, 1260, and 567, respectively. In most cases, limiting the analyses to participants who reached a certain age had a minimal impact on both the regression coefficients and the evaluation of a model's performance. Apparent performance measures remained mostly the same, while optimism-corrected performance measures tended to gradually decrease as the age restriction increased (Online supplement).

## **DISCUSSION**

We developed and internally validated eight clinical risk scores that estimate the risk of eventually developing AML or LAM using clinical TSC features seen during childhood. All models showed good to strong capacity to discriminate between people who will develop the associated outcome and those who will not. Calibration was strong in all AML models and good in LAM models (deviations seen at predicted risks greater than 0.2). More parsimonious models using the LASSO method resulted in minimal losses in predictive performance.

Prior studies aiming to identify predictors of AML or LAM published in the literature are limited to descriptive analyses comparing outcome frequency according to certain clinical features such as biological sex, age, and genetic mutations [3, 7]. Our models confirmed both sex and TSC1/TSC2 genetic mutations as strong predictive factors, while also identifying several other predictors among the clinical manifestations of TSC seen before the age of five.

The development of clinical risk scores able to accurately predict the risk of developing AML or LAM has important implications in both the screening and treatment of patients.

Imaging and screening schedules aiming to detect the presence or to follow the course of renal and pulmonary tumours are currently not adapted to a patient's individual risk and don't take into consideration the significant heterogeneity of TSC. Our findings are a significant step towards more tailored schedules for screening and imaging in people with TSC. There is increasing interest in modifying the course of the disease for people affected by TSC. Prophylactic use of the antiseizure medication vigabatrin in infants during the EPISTOP trial was shown to delay the first clinical seizures as well as decreasing the risk of developing drug resistant epilepsy [61]. The use

of mTOR inhibitors (e.g., everolimus and sirolimus) has already been established as effective in treating both AML lesions and LAM in randomised clinical trials [12, 18]. We expect our clinical risk scores will become essential in identifying individuals with TSC who would most likely benefit from prophylactic or disease-modifying therapies, whether with antiseizure medications or mTOR inhibitors. This will prove useful in clinical practice as well as in the design of future preventative trials in TSC.

Our study has several strengths. Our models were developed using the TSC-NHD, the largest longitudinal cohort of people with TSC reported in the literature. Three of the most important aspects of predictive performance were assessed, as recommended by the TRIPOD guidelines [20]. Predictive performance of the models was further confirmed by bootstrapping, a robust form of internal validation.

Our study has limitations. Recruitment of participants in the TSC-NHD was neither consecutive nor random but was carried out using a convenience sampling method. The use of a non-probabilistic sampling method is known to potentially cause selection bias, in which certain subgroups of the population have a higher chance of being included in the study sample than others, for reasons related to the exposure and outcome. Even in the absence of selection bias, this non-probabilistic sampling can lead to a lack of representativeness of the study population, which in turn may lead to difficulties when attempting to generalise estimates to all patients with TSC. An important amount of missing data was identified in some of the predictors selected in the development of our models. Even though multiple imputation by chained equations is an efficient method to replace missing data, the possibility of biased estimates resulting from the missing information cannot be excluded. External validation of our risk scores in a separate large cohort of people with TSC will be an important next step to confirming the robustness of our findings.

We developed and internally validated a series of clinical risk scores that aim to predict incident renal AML and pulmonary LAM in people with TSC. Our findings serve as an important first step toward assuring more personalised medicine for people with TSC.

### **Figure legends**

**Figure 1.** Calibration plots comparing the predicted probabilities to the ones observed in the sample were evaluated for each of the four outcomes. Every model's calibration is showed by the blue line, while the black line with a slope of 1 denotes perfect calibration.

**Figure 2.** The horizontal axis refers to the clinical risk score calculated using Table 3, while the vertical axis refers to the risk for an individual to develop the corresponding outcome.

## REFERENCES

1. Henske EP, Jozwiak S, Kingswood JC, Sampson JR, Thiele EA. Tuberous sclerosis complex. *Nat Rev Dis Primers*. 2016;2:16035.
2. Lyczkowski DA, Conant KD, Pulsifer MB, Jarrett DY, Grant PE, Kwiatkowski DJ, et al. Intrafamilial phenotypic variability in tuberous sclerosis complex. *J Child Neurol*. 2007;22(12):1348-55.
3. Rakowski SK, Winterkorn EB, Paul E, Steele DJ, Halpern EF, Thiele EA. Renal manifestations of tuberous sclerosis complex: Incidence, prognosis, and predictive factors. *Kidney Int*. 2006;70(10):1777-82.
4. Cudzilo CJ, Szczesniak RD, Brody AS, Rattan MS, Krueger DA, Bissler JJ, et al. Lymphangiomyomatosis screening in women with tuberous sclerosis. *Chest*. 2013;144(2):578-85.
5. Rosset C, Netto CBO, Ashton-Prolla P. TSC1 and TSC2 gene mutations and their implications for treatment in Tuberous Sclerosis Complex: a review. *Genet Mol Biol*. 2017;40(1):69-79.
6. Northrup H, Aronow ME, Bebin EM, Bissler J, Darling TN, de Vries PJ, et al. Updated International Tuberous Sclerosis Complex Diagnostic Criteria and Surveillance and Management Recommendations. *Pediatr Neurol*. 2021;123:50-66.
7. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
8. Yamakado K, Tanaka N, Nakagawa T, Kobayashi S, Yanagawa M, Takeda K. Renal angiomyolipoma: relationships between tumor size, aneurysm formation, and rupture. *Radiology*. 2002;225(1):78-82.
9. Kingswood JC, d'Augeres GB, Belousova E, Ferreira JC, Carter T, Castellana R, et al. Tuberous Sclerosis registry to increase disease Awareness (TOSCA) - baseline data on 2093 patients. *Orphanet J Rare Dis*. 2017;12(1):2.
10. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *Journal of statistical software*. 2011;45:1-67.
11. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine*. 2011;30(4):377-99.
12. Moons KG, Donders RA, Stijnen T, Harrell Jr FE. Using the outcome for imputation of missing predictor values was preferred. *Journal of clinical epidemiology*. 2006;59(10):1092-101.
13. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.
14. Vergouwe Y, Royston P, Moons KG, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. *Journal of clinical epidemiology*. 2010;63(2):205-14.
15. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological)*. 1996;58(1):267-88.
16. Bonnett LJ, Snell KIE, Collins GS, Riley RD. Guide to presenting clinical prediction models for use in clinical settings. *BMJ*. 2019;365:l737.

17. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1-73.
18. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* 2001;54(8):774-81.
19. Steyerberg EW. Overfitting and optimism in prediction models. *Clinical prediction models*: Springer; 2019. p. 95-112.
20. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 2009;338:b605.
21. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass).* 2010;21(1):128.
22. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*: John Wiley & Sons; 2013.
23. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software.* 2011;45(3):1 - 67.
24. Robitzsch A, Grund S. miceadds: Some Additional Multiple Imputation Functions, Especially for 'mice'. R package version 3.16-18 ed2023.
25. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1-22.
26. Harrell Jr FE. rms: Regression Modeling Strategies. R package version 6.3-0 ed2022.
27. Kotulska K, Kwiatkowski DJ, Curatolo P, Weschke B, Riney K, Jansen F, et al. Prevention of Epilepsy in Infants with Tuberous Sclerosis Complex in the EPISTOP Trial. *Ann Neurol.* 2021;89(2):304-14.
28. Bissler JJ, Kingswood JC, Radzikowska E, Zonnenberg BA, Frost M, Belousova E, et al. Everolimus for angiomyolipoma associated with tuberous sclerosis complex or sporadic lymphangioleiomyomatosis (EXIST-2): a multicentre, randomised, double-blind, placebo-controlled trial. *Lancet.* 2013;381(9869):817-24.
29. McCormack FX, Inoue Y, Moss J, Singer LG, Strange C, Nakata K, et al. Efficacy and safety of sirolimus in lymphangioleiomyomatosis. *N Engl J Med.* 2011;364(17):1595-606.

## 5.2. Tableaux

**Table 1. Descriptive Statistics for Participants**

	AML analyses (n=1939)			LAM analyses (n=2420)		
	n (%) *					
Characteristics	Value		Missing	Value		Missing
<b>Biological sex – Female</b>	967 (49.87)		1 (0.05)	1210 (50.00)		2 (0.08)
<b>Median follow-up time</b>	7.07		-	6.78		-
<b>Age at last data entry</b>			-			22 (0.91)
<= 12	448 (23.10)		-	738 (30.50)		-
13-18	389 (20.06)		-	460 (19.01)		-
19-30	549 (28.31)		-	633 (26.16)		-
>30	553 (28.52)		-	567 (23.43)		-
<b>Molecular testing</b>						-
TSC1	298 (15.37)		-	330 (13.64)		-
TSC2	587 (30.27)		-	840 (34.71)		-
Unknown/unmeasured	1054 (54.36)		-	1250 (51.65)		-
<b>Clinical assessment of IQ</b>			-			-
Normal intellectual ability	260 (13.41)		-	291 (12.02)		-
Mild-moderate intellectual disability	398 (20.53)		-	497 (20.54)		-
Severe-profound intellectual disability	143 (7.37)		-	184 (7.60)		-
Unknown/unmeasured	1138 (58.69)		-	1448 (59.83)		-
<b>Clinical manifestations</b>	<b>Present</b>	<b>Absent</b>	<b>Missing</b>	<b>Present</b>	<b>Absent</b>	<b>Missing</b>
Subependymal nodules	1522 (78.49)	247 (12.74)	170 (8.77)	1939 (80.12)	268 (11.07)	213 (8.80)
Cortical tuber hamartomas	1565 (80.71)	215 (11.09)	159 (8.20)	1965 (81.20)	259 (10.70)	196 (8.10)
Subependymal giant cell astrocytoma (SEGA)	524 (27.02)	1226 (63.23)	189 (9.75)	675 (27.89)	1508 (62.31)	237 (9.79)
Epilepsy	1566 (80.76)	283 (14.60)	90 (4.64)	1976 (81.65)	322 (13.31)	122 (5.04)
Facial angiofibromas	1258 (64.88)	539 (27.80)	142 (7.32)	1510 (62.40)	720 (29.75)	190 (7.85)



Hypomelanotic macules (3 or more)	1320 (68.08)	376 (19.39)	243 (12.53)	1685 (69.63)	441 (18.22)	294 (12.15)
Shagreen patch	693 (35.74)	1043 (53.79)	203 (10.47)	850 (35.12)	1314 (54.30)	256 (10.58)
Cardiac rhabdomyoma	759 (39.14)	912 (47.03)	268 (13.82)	1081 (44.67)	1025 (42.36)	314 (12.98)
Retinal hamartoma	306 (15.78)	1283 (66.17)	350 (18.05)	417 (17.23)	1574 (65.04)	429 (17.73)
AML	889 (45.85)	867 (44.71)	183 (9.44)	1178 (48.68)	1007 (41.61)	235 (9.71)
Bilateral AML	633 (32.65)	1040 (53.64)	266 (13.72)	847 (35.00)	1239 (51.20)	334 (13.80)
AML with lesions over 3cm	276 (14.23)	1336 (68.90)	327 (16.86)	325 (13.43)	1680 (69.42)	415 (17.15)
LAM	101 (5.21)	1838 (94.79)	-	107 (4.42)	2313 (95.58)	-

\*Except for median follow-up time which refers to the median number of years of follow-up.

**Table 2. Full Regression Models**

	<b>AML</b>	<b>Bilateral AML</b>	<b>AML lesions greater than 3cm</b>	<b>LAM</b>
<b>Variables</b>	Odds ratio (95% CI)			
<b>Biological sex - Women</b>	1.55 (1.23 – 1.96)	1.63 (1.27 – 2.09)	1.73 (1.36 – 2.25)	21.89 (8.75 – 54.74)
<b>TSC2 mutation</b>	2.58 (1.95 - 3.41)	3.80 (2.79 – 5.17)	3.80 (2.13 – 6.78)	2.75 (1.32 – 5.72)
<b>Moderate intellectual disability</b>	1.48 (1.09 – 2.01)	1.35 (0.92 – 1.97)	1.04 (0.66 – 1.61)	0.69 (0.37 – 1.29)
<b>Severe intellectual disability</b>	1.80 (1.20 – 2.72)	1.82 (1.16 – 2.85)	1.30 (0.82 – 2.05)	0.90 (0.40 – 2.03)
<b>Subependymal nodules</b>	1.60 (1.08 – 2.35)	1.58 (1.05 – 2.39)	1.50 (0.86 – 2.59)	1.12 (0.51 – 2.48)
<b>Cortical tubers</b>	0.76 (0.52 – 1.10)	0.89 (0.59 – 1.36)	0.78 (0.44 – 1.40)	0.66 (0.31 – 1.42)
<b>Subependymal giant cell astrocytoma (SEGA)</b>	2.05 (1.56 – 2.68)	2.04 (1.58 – 2.65)	1.29 (0.95 – 1.76)	0.88 (0.52 – 1.49)
<b>Epilepsy</b>	0.72 (0.49 – 1.05)	0.86 (0.56 – 1.34)	0.71 (0.42 – 1.18)	0.45 (0.23 – 0.89)
<b>Facial angiofibroma</b>	5.17 (3.95 – 6.77)	5.07 (3.67 – 7.02)	4.02 (2.62 – 6.15)	3.02 (1.55 – 5.87)
<b>Hypomelanotic macules</b>	0.61 (0.45 – 0.82)	0.69 (0.50 – 0.96)	0.56 (0.39 – 0.81)	0.60 (0.34 – 1.06)
<b>Shagreen patch</b>	1.60 (1.27 – 2.02)	1.62 (1.27 – 2.06)	1.66 (1.25 – 2.21)	1.22 (0.74 – 2.01)
<b>Cardiac rhabdomyoma</b>	0.81 (0.63 – 1.04)	0.70 (0.54 – 0.90)	0.68 (0.49 – 0.95)	0.30 (0.14 – 0.64)
<b>Retinal hamartoma</b>	2.36 (1.74 – 3.19)	1.94 (1.46 – 2.59)	1.70 (1.17 – 2.47)	1.71 (0.93 – 3.15)
<b>Apparent C-Index (95% CI)</b>	0.79 (0.78 – 0.81)	0.80 (0.78 – 0.82)	0.76 (0.74 – 0.79)	0.86 (0.84 – 0.90)
<b>Optimism-corrected C-Index (95% CI)</b>	0.79 (0.77 – 0.81)	0.79 (0.78 – 0.81)	0.75 (0.73 – 0.78)	0.85 (0.82 – 0.89)

**Table 3. Clinical Risk Scores\* for Full Regression Models**

	<b>AML</b>	<b>Bilateral AML</b>	<b>AML over 3cm</b>	<b>LAM</b>
<b>Biological sex - Female</b>	3.1	3.2	4.7	22.4
<b>TSC2 mutation</b>	6.8	8.8	11.5	7.3
<b>Moderate intellectual disability</b>	2.8	1.9	0.3	-2.7
<b>Severe intellectual disability</b>	4.2	3.9	2.2	-0.8
<b>Subependymal nodules</b>	3.3	3.0	3.5	0.8
<b>Cortical tubers</b>	-2.0	-0.7	-2.1	-3.0
<b>Subependymal giant cell astrocytoma (SEGA)</b>	5.1	4.7	2.2	-0.9
<b>Epilepsy</b>	-2.3	-1.0	-3.0	-5.8
<b>Facial angiofibroma</b>	11.7	10.6	12.0	8.0
<b>Hypomelanotic macules</b>	-3.6	-2.4	-5.0	-3.8
<b>Shagreen patch</b>	3.4	3.2	4.4	1.4
<b>Cardiac rhabdomyoma</b>	-1.5	-2.4	-3.3	-8.7
<b>Retinal hamartoma</b>	6.1	4.4	4.6	3.9

\*Regression coefficients were converted into clinical risk scores by applying weights to the regression coefficients, allowing us to evaluate an individual's risk when used in combination with Table 4.

**Table 4. Probability (95% CI) of developing the outcome associated with the total score\***

<b>Total score</b>	<b>AML</b>	<b>Bilateral AML</b>	<b>AML over 3cm</b>	<b>LAM</b>
<b>-10</b>	-	-	0.01 (0.01 – 0.02)	0.00 (0.00 – 0.00)
<b>-5</b>	0.05 (0.03 – 0.07)	0.02 (0.01 – 0.03)	0.02 (0.01 – 0.03)	0.00 (0.00 – 0.01)
<b>0</b>	0.10 (0.07 – 0.15)	0.04 (0.02 – 0.06)	0.03 (0.02 – 0.05)	0.00 (0.00 – 0.01)
<b>5</b>	0.19 (0.13 – 0.25)	0.08 (0.05 – 0.12)	0.05 (0.03 – 0.08)	0.01 (0.00 – 0.02)
<b>10</b>	0.31 (0.24 – 0.40)	0.15 (0.10 – 0.21)	0.08 (0.05 – 0.13)	0.02 (0.01 – 0.04)
<b>15</b>	0.47 (0.38 – 0.56)	0.26 (0.19 – 0.35)	0.14 (0.09 – 0.21)	0.03 (0.01 – 0.07)
<b>20</b>	0.64 (0.55 – 0.71)	0.43 (0.33 – 0.52)	0.22 (0.15 – 0.32)	0.06 (0.03 – 0.12)
<b>25</b>	0.78 (0.71 – 0.83)	0.61 (0.52 – 0.69)	0.33 (0.24 – 0.42)	0.12 (0.06 – 0.21)
<b>30</b>	0.87 (0.82 – 0.91)	0.76 (0.68 – 0.82)	0.46 (0.36 – 0.57)	0.22 (0.12 – 0.37)
<b>35</b>	0.93 (0.89 – 0.96)	0.87 (0.81 – 0.91)	0.60 (0.49 – 0.70)	0.37 (0.21 – 0.55)
<b>40</b>	0.96 (0.94 – 0.98)	0.93 (0.89 – 0.96)	0.72 (0.61 – 0.82)	0.54 (0.33 – 0.74)

\*Total score is evaluated using Table 3 based on an individual's TSC features, along with their sex and genetic mutation.

### 5.3. Figures

Figure 1. Calibration plots.

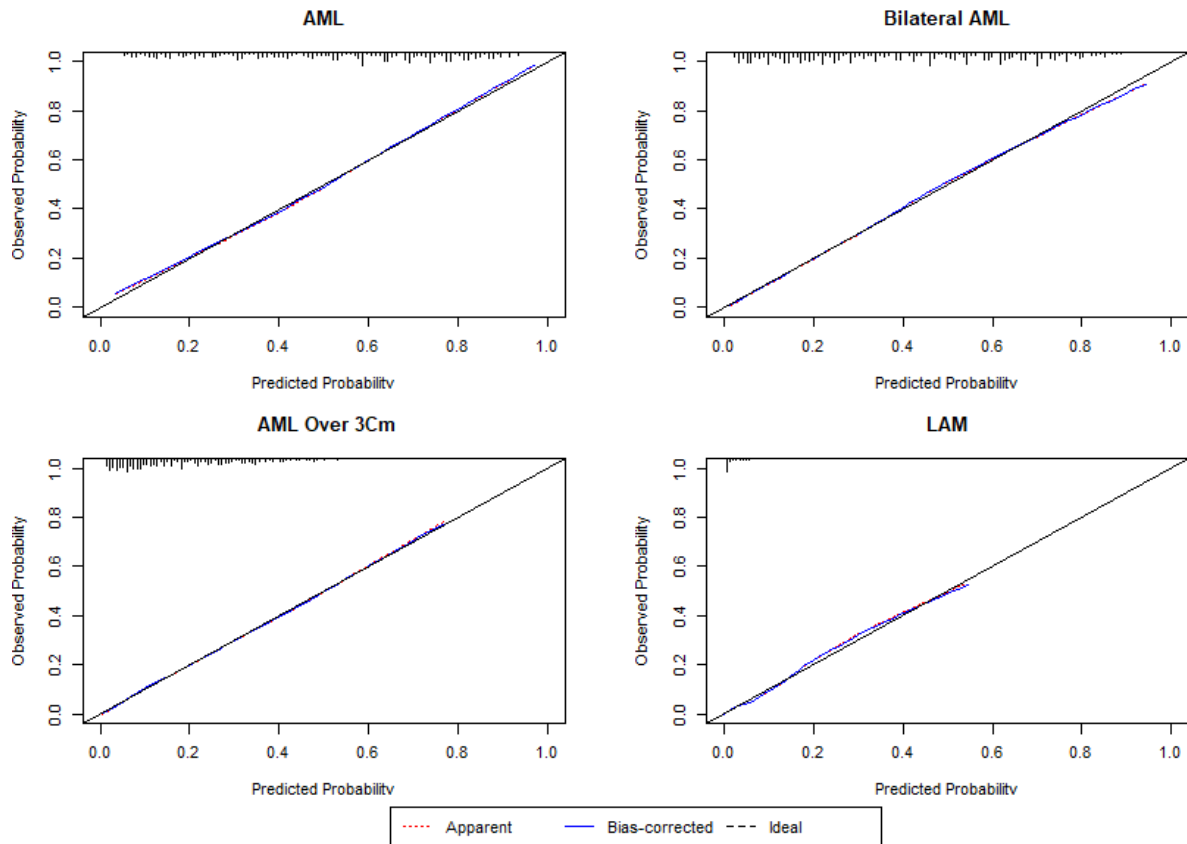
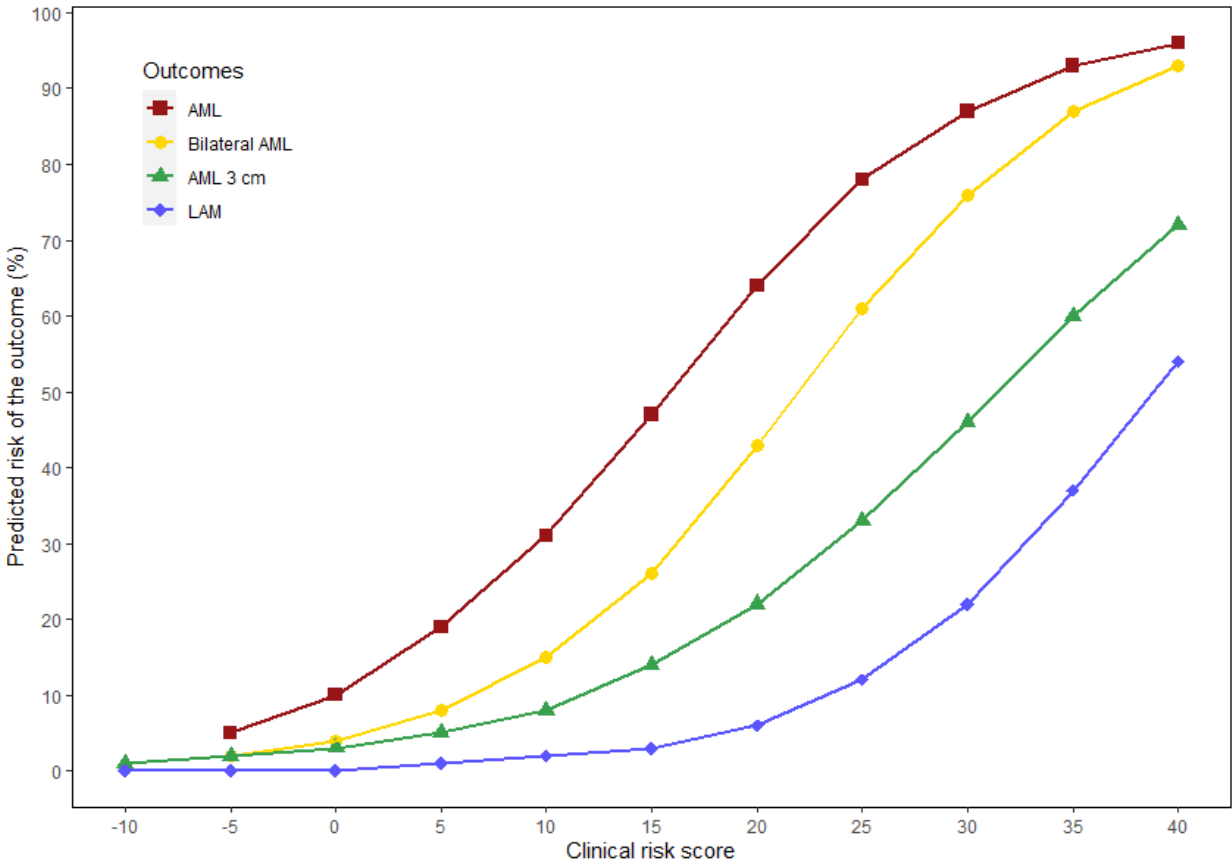


Figure 2. Clinical risk percentage by total score



## 5.4. Suppléments en ligne

### Online supplement 1. LASSO models

	AML	Bilateral AML	AML lesions greater than 3cm	LAM
<b>Variables</b>	<b>Odds ratio</b>			
Regression model intercept	0.16	0.07	0.05	0.02
Biological sex - Women	1.19	1.22	1.27	4.68
TSC2 mutation	2.13	2.86	2.42	1.14
Moderate intellectual disability	-	-	-	-
Severe intellectual disability	1.10	1.21	-	-
Subependymal nodules	1.04	1.06	-	-
Cortical tubers	-	-	-	-
Subependymal giant cell astrocytoma (SEGA)	1.65	1.64	1.03	-
Epilepsy	-	-	-	0.63
Facial angiofibroma	3.93	3.71	2.58	1.25
Hypomelanotic macules	0.85	-	0.85	-
Shagreen patch	1.29	1.30	1.26	-
Cardiac rhabdomyoma	-	0.95	0.92	0.56
Retinal hamartoma	1.85	1.58	1.37	-
<b>Apparent C-Index</b>	0.78	0.79	0.75	0.85
<b>Optimism-corrected C-Index</b>	0.78	0.79	0.74	0.84

**Online supplement 2. Clinical risk score by predictor for LASSO models**

	<b>AML</b>	<b>Bilateral AML</b>	<b>AML over 3cm</b>	<b>LAM</b>
<b>Biological sex - Female</b>	1.9	2.0	3.6	32.0
<b>TSC2 mutation</b>	7.9	10.5	13.3	3.5
<b>Moderate intellectual disability</b>	-	-	-	-
<b>Severe intellectual disability</b>	1.1	1.9	-	-
<b>Subependymal nodules</b>	0.6	0.6	-	-
<b>Cortical tubers</b>	-	-	-	-
<b>Subependymal giant cell astrocytoma (SEGA)</b>	5.2	5.0	0.4	-
<b>Epilepsy</b>	-	-	-	-10
<b>Facial angiofibroma</b>	14.2	13.1	14.3	5.2
<b>Hypomelanotic macules</b>	-1.9	-	-2.5	-
<b>Shagreen patch</b>	2.8	2.6	3.5	-
<b>Cardiac rhabdomyoma</b>	0.0	-0.5	-1.2	-12.3
<b>Retinal hamartoma</b>	6.4	4.6	4.8	-



**Online supplement 3. Probability (95% CI) of outcome during follow-up by total clinical risk score for LASSO models**

<b>Total score</b>	<b>AML</b>	<b>Bilateral AML</b>	<b>AML over 3cm</b>	<b>LAM</b>
<b>-10</b>	-	-	-	0·00 (0·00 – 0·01)
<b>-5</b>	-	-	0·01 (0·01 – 0·02)	0·00 (0·00 – 0·01)
<b>0</b>	0·08 (0·05 – 0·13)	0·03 (0·02 – 0·05)	0·02 (0·01 – 0·03)	0·01 (0·00 – 0·02)
<b>5</b>	0·14 (0·09 – 0·21)	0·06 (0·04 – 0·09)	0·03 (0·02 – 0·05)	0·01 (0·00 – 0·03)
<b>10</b>	0·24 (0·17 – 0·32)	0·11 (0·07 – 0·17)	0·05 (0·03 – 0·08)	0·02 (0·01 – 0·04)
<b>15</b>	0·37 (0·28 – 0·46)	0·19 (0·13 – 0·27)	0·08 (0·05 – 0·12)	0·03 (0·01 – 0·06)
<b>20</b>	0·52 (0·41 – 0·62)	0·31 (0·22 – 0·41)	0·13 (0·08 – 0·20)	0·05 (0·02 – 0·11)
<b>25</b>	0·67 (0·57 – 0·75)	0·47 (0·35 – 0·58)	0·20 (0·14 – 0·28)	0·07 (0·03 – 0·16)
<b>30</b>	0·79 (0·72 – 0·84)	0·63 (0·54 – 0·71)	0·30 (0·21 – 0·40)	0·12 (0·06 – 0·21)
<b>35</b>	0·87 (0·82 – 0·91)	0·77 (0·69 – 0·83)	0·42 (0·30 – 0·54)	0·18 (0·09 – 0·32)
<b>40</b>	0·93 (0·88 – 0·96)	0·86 (0·81 – 0·91)	0·55 (0·42 – 0·68)	0·27 (0·15 – 0·43)

### Online supplement 4. AML sensitivity analysis

Predictors	Age at last follow-up							
	All (n= 1939)		> 12 (n=1564)		> 18 (n= 1153)		> 30 (n= 553)	
	Odds ratio (95% CI)							
AML n – (%)	1000 (51.57)		906 (57.93)		732 (63.49)		382 (69.08)	
Intercept	0.11 (0.06 - 0.18)		0.11 (0.07 - 0.20)		0.12 (0.06 - 0.24)		0.21 (0.08 - 0.52)	
Biological sex – Women	1.55 (1.23 - 1.96)		1.77 (1.37 - 2.28)		1.96 (1.45 - 2.65)		1.91 (1.18 - 3.09)	
TSC2 mutation	2.58 (1.95 - 3.41)		2.89 (2.11 - 3.93)		2.85 (1.96 - 4.14)		3.79 (1.97 - 7.27)	
Moderate intellectual disability	1.48 (1.09 - 2.01)		1.59 (1.09 - 2.33)		1.55 (0.99 - 2.42)		1.41 (0.76 - 2.61)	
Severe intellectual disability	1.80 (1.20 - 2.72)		1.90 (1.17 - 3.07)		1.76 (1.00 - 3.11)		2.19 (0.84 - 5.74)	
Subependymal nodules	1.60 (1.08 - 2.35)		1.87 (1.23 - 2.85)		2.23 (1.39 - 3.56)		2.11 (1.11 - 3.98)	
Cortical tubers	0.76 (0.52 - 1.10)		0.81 (0.53 - 1.23)		0.68 (0.42 - 1.10)		0.78 (0.40 - 1.53)	
SEGA	2.05 (1.56 - 2.68)		1.90 (1.41 - 2.55)		1.63 (1.14 - 2.33)		1.23 (0.68 - 2.19)	
Epilepsy	0.72 (0.49 - 1.05)		0.59 (0.39 - 0.91)		0.66 (0.41 - 1.07)		0.51 (0.26 - 0.99)	
Facial angiofibroma	5.17 (3.95 - 6.77)		3.93 (2.87 - 5.37)		3.38 (2.30 - 4.98)		2.49 (1.39 - 4.45)	
Hypomelanotic macules	0.61 (0.45 - 0.82)		0.64 (0.45 - 0.89)		0.80 (0.56 - 1.16)		0.87 (0.53 - 1.45)	
Shagreen patch	1.60 (1.27 - 2.02)		1.47 (1.13 - 1.90)		1.26 (0.93 - 1.70)		1.32 (0.81 - 2.14)	
Cardiac rhabdomyoma	0.81 (0.63 - 1.04)		1.05 (0.78 - 1.42)		1.27 (0.87 - 1.85)		1.02 (0.51 - 2.04)	
Retinal hamartoma	2.36 (1.74 - 3.19)		2.60 (1.81 - 3.72)		3.13 (1.96 - 5.01)		3.96 (1.57 - 9.94)	
	Performance							
	App.	Corr.	App.	Corr.	App.	Corr.	App.	Corr.
Nagelkerke's R <sup>2</sup>	0.33	0.31	0.31	0.29	0.29	0.27	0.29	0.24
C-index	0.79	0.79	0.79	0.78	0.78	0.77	0.79	0.76

**Online supplement 5. Bilateral AML sensitivity analysis**

Predictors	Age at last follow-up							
	All (n= 1939)		> 12 (n=1564)		> 18 (n= 1153)		> 30 (n= 553)	
	Odds ratio (95% CI)							
<b>Bilateral AML n – (%)</b>	792 (40-85)		730 (46-68)		591 (51-26)		309 (55-88)	
<b>Intercept</b>	0-03 (0-02 – 0-06)		0-04 (0-02 – 0-07)		0-04 (0-02 – 0-09)		0-08 (0-03 – 0-24)	
<b>Biological sex – Women</b>	1-63 (1-27 – 2-09)		1-76 (1-35 – 2-28)		1-82 (1-35 – 2-46)		1-53 (0-95 – 2-47)	
<b>TSC2 mutation</b>	3-80 (2-79 – 5-17)		4-25 (2-97 – 6-08)		4-39 (2-94 – 6-56)		5-64 (2-90 – 10-99)	
<b>Moderate intellectual disability</b>	1-35 (0-92 – 1-97)		1-41 (0-90 – 2-23)		1-42 (0-86 – 2-33)		1-61 (0-86 – 3-01)	
<b>Severe intellectual disability</b>	1-82 (1-16 – 2-85)		1-82 (1-09 – 3-02)		1-67 (1-01 – 2-77)		1-92 (0-91 – 4-05)	
<b>Subependymal nodules</b>	1-58 (1-05 – 2-39)		1-81 (1-16 – 2-81)		2-05 (1-25 – 3-37)		1-57 (0-83 – 2-99)	
<b>Cortical tubers</b>	0-89 (0-59 – 1-36)		0-95 (0-60 – 1-51)		0-83 (0-48 – 1-42)		1-03 (0-51 – 2-07)	
<b>SEGA</b>	2-04 (1-58 – 2-65)		1-97 (1-50 – 2-59)		1-77 (1-27 – 2-45)		1-64 (0-98 – 2-76)	
<b>Epilepsy</b>	0-86 (0-56 – 1-34)		0-73 (0-45 – 1-18)		0-77 (0-47 – 1-26)		0-57 (0-30 – 1-09)	
<b>Facial angiofibroma</b>	5-07 (3-67 – 7-02)		3-76 (2-62 – 5-41)		3-28 (2-13 – 5-05)		2-37 (1-30 – 4-31)	
<b>Hypomelanotic macules</b>	0-69 (0-50 – 0-96)		0-74 (0-45 – 1-08)		0-97 (0-65 – 1-43)		1-01 (0-63 – 1-63)	
<b>Shagreen patch</b>	1-62 (1-27 – 2-06)		1-44 (1-11 – 1-88)		1-22 (0-90 – 1-65)		1-16 (0-72 – 1-86)	
<b>Cardiac rhabdomyoma</b>	0-70 (0-54 – 0-90)		0-84 (0-63 – 1-13)		0-90 (0-62 – 1-31)		0-67 (0-34 – 1-33)	
<b>Retinal hamartoma</b>	1-94 (1-46 – 2-59)		2-06 (1-49 – 2-85)		2-22 (1-47 – 3-35)		1-81 (0-95 – 3-44)	
	Performance							
	App.	Corr.	App.	Corr.	App.	Corr.	App.	Corr.
<b>Nagelkerke’s R<sup>2</sup></b>	0-34	0-32	0-32	0-30	0-31	0-29	0-30	0-25
<b>C-index</b>	0-80	0-79	0-79	0-78	0-78	0-77	0-78	0-75

**Online supplement 6. AML with lesions > 3cm sensitivity analysis**

Predictors	Age at last follow-up							
	All (n= 1939)		> 12 (n=1564)		> 18 (n= 1153)		> 30 (n= 553)	
	Odds ratio (95% CI)							
AML >3cm n – (%)	378 (19.49)		360 (23.02)		326 (28.27)		204 (36.89)	
Intercept	0.03 (0.01 – 0.05)		0.03 (0.02 -0.07)		0.03 (0.02 – 0.08)		0.05 (0.02 – 0.14)	
Biological sex – Women	1.73 (1.36 – 2.25)		1.78 (1.34 – 2.38)		1.86 (1.36 – 2.54)		1.63 (1.05 – 2.54)	
TSC2 mutation	3.80 (2.13 – 6.78)		4.00 (2.21 – 7.26)		4.23 (2.32 – 7.73)		5.27 (2.22 – 12.48)	
Moderate intellectual disability	1.04 (0.66 – 1.61)		1.09 (0.67 – 1.78)		1.10 (0.63 – 1.92)		1.06 (0.52 – 2.15)	
Severe intellectual disability	1.30 (0.82 – 2.05)		1.32 (0.81 – 2.15)		1.16 (0.68 – 1.98)		1.16 (0.57 – 2.38)	
Subependymal nodules	1.50 (0.86 – 2.59)		1.69 (0.98 – 2.92)		2.00 (1.06 – 3.75)		1.89 (0.90 – 3.96)	
Cortical tubers	0.78 (0.44 – 1.40)		0.76 (0.42 – 1.38)		0.66 (0.36 – 1.22)		0.79 (0.34 – 1.83)	
SEGA	1.29 (0.95 – 1.76)		1.26 (0.91 – 1.73)		1.10 (0.77 – 1.56)		1.30 (0.78 – 2.15)	
Epilepsy	0.71 (0.42 – 1.18)		0.65 (0.38 – 1.13)		0.72 (0.40 – 1.30)		0.58 (0.26 – 1.26)	
Facial angiofibroma	4.02 (2.62 – 6.15)		2.85 (1.82 – 4.47)		2.68 (1.60 – 4.48)		2.41 (1.12 – 5.20)	
Hypomelanotic macules	0.56 (0.39 – 0.81)		0.57 (0.38 – 0.84)		0.63 (0.42 – 0.94)		0.70 (0.42 – 1.15)	
Shagreen patch	1.66 (1.25 – 2.21)		1.61 (1.21 – 2.15)		1.64 (1.17 – 2.28)		1.59 (0.96 – 2.64)	
Cardiac rhabdomyoma	0.68 (0.49 – 0.95)		0.83 (0.59 – 1.16)		0.93 (0.64 – 1.37)		0.86 (0.47 – 1.59)	
Retinal hamartoma	1.70 (1.17 – 2.47)		1.77 (1.19 – 2.63)		1.75 (1.14 – 2.69)		2.27 (1.24 – 4.16)	
	Performance							
	App.	Corr.	App.	Corr.	App.	Corr.	App.	Corr.
Nagelkerke's R <sup>2</sup>	0.21	0.19	0.19	0.17	0.20	0.17	0.25	0.20
C-index	0.76	0.75	0.74	0.73	0.74	0.72	0.76	0.73

## Online supplement 7. LAM sensitivity analysis

Predictors	Age at last follow-up							
	All (n= 2420)		> 12 (n=1750)		> 18 (n= 1260)		> 30 (n= 567)	
	Odds ratio (95% CI)							
LAM n – (%)	107 (4.42)	107 (6.11)	106 (8.41)	89 (15.70)				
Intercept	0.00 (0.00 – 0.02)	0.01 (0.00 – 0.02)	0.01 (0.00 – 0.03)	0.01 (0.00 – 0.04)				
Biological sex – Women	21.89 (8.75 – 54.74)	21.99 (8.77 – 55.15)	21.87 (8.68 – 55.11)	19.66 (7.45 – 51.90)				
TSC2 mutation	2.75 (1.32 – 5.72)	2.88 (1.39 – 5.96)	2.88 (1.35 – 6.15)	4.14 (1.73 – 9.92)				
Moderate intellectual disability	0.69 (0.37 – 1.29)	0.71 (0.38 – 1.32)	0.68 (0.36 – 1.28)	0.66 (0.30 – 1.46)				
Severe intellectual disability	0.90 (0.40 – 2.03)	0.90 (0.40 – 2.05)	0.83 (0.37 – 1.88)	0.72 (0.23 – 2.21)				
Subependymal nodules	1.12 (0.51 – 2.48)	1.22 (0.53 – 2.81)	1.40 (0.58 – 3.37)	1.48 (0.54 – 4.03)				
Cortical tubers	0.66 (0.31 – 1.42)	0.62 (0.28 – 1.36)	0.61 (0.26 – 1.41)	0.82 (0.34 – 1.97)				
SEGA	0.88 (0.52 – 1.49)	0.84 (0.49 – 1.43)	0.89 (0.52 – 1.53)	0.64 (0.32 – 1.29)				
Epilepsy	0.45 (0.23 – 0.89)	0.43 (0.21 – 0.87)	0.45 (0.22 – 0.92)	0.46 (0.19 – 1.07)				
Facial angiofibroma	3.02 (1.55 – 5.87)	2.09 (1.04 – 4.20)	1.72 (0.82 – 3.63)	1.29 (0.57 – 2.93)				
Hypomelanotic macules	0.60 (0.34 – 1.06)	0.65 (0.36 – 1.18)	0.74 (0.41 – 1.36)	1.03 (0.52 – 2.07)				
Shagreen patch	1.22 (0.74 – 2.01)	1.10 (0.66 – 1.83)	1.05 (0.62 – 1.78)	0.83 (0.43 – 1.60)				
Cardiac rhabdomyoma	0.30 (0.14 – 0.64)	0.43 (0.20 – 0.90)	0.55 (0.25 – 1.20)	0.52 (0.15 – 1.75)				
Retinal hamartoma	1.71 (0.93 – 3.15)	1.72 (0.93 – 3.18)	1.73 (0.92 – 3.26)	2.64 (1.16 – 6.01)				
	Performance							
	App.	Corr.	App.	Corr.	App.	Corr.	App.	Corr.
Nagelkerke's R <sup>2</sup>	0.27	0.24	0.26	0.23	0.26	0.23	0.34	0.28
C-index	0.86	0.85	0.85	0.83	0.83	0.81	0.84	0.81

## **5.5. Contribution des auteurs**

- Frédéric Loubert: Collecte des données, conception de l'étude, analyses des données, écriture du manuscrit, révision et correction du manuscrit.
- Mark R. Keezer: Conceptualisation du projet de recherche, supervision du projet de recherche, conception de l'étude, collecte des données, révision et correction de l'analyse des données, révision et correction du manuscrit.
- Andrew A. House: Conception de l'étude, révision et correction du manuscrit.
- Catherine Larochelle: Conception de l'étude, révision et correction du manuscrit.
- Philippe Major: Conception de l'étude, révision et correction du manuscrit.

## Chapitre 6 — Discussion

### 6.1. Sommaire des résultats

Le présent projet de recherche constitue une étude visant la création de scores de risque clinique servant à prédire l'apparition d'angiomyolipomes rénaux et de lymphangioléiomyomatose pulmonaire chez les personnes atteintes de la STB. L'hypothèse initiale guidant l'objectif principal de cette étude était que les manifestations cliniques de la STB qui surviennent avant l'âge de 5 ans peuvent être utilisées afin de prédire le risque futur d'AML et de LAM. Les résultats obtenus par l'élaboration de modèles de régression logistique adaptés en scores de risque ont permis de confirmer cette hypothèse. Les modèles développés sont en mesure de prédire de façon adéquate la probabilité qu'une personne contracte éventuellement l'une des issues étudiées. Ceux-ci sont d'ailleurs simples à employer dans un contexte clinique et ne demandent que de l'information au sujet de manifestations cliniques répandues et dont le diagnostic est généralement connu par les cliniciens effectuant le suivi de leurs patients atteints de la STB.

Il est important de noter avant d'entrer plus en détail dans le sommaire des résultats qu'étant donné que l'objectif n'était pas d'acquérir une meilleure compréhension de l'étiologie des issues étudiées, les liens qui les unissent aux prédicteurs ne sont que corrélationnels et ne devraient donc pas être interprétés de manière causale. À cet effet, les modèles développés sont évalués dans leur ensemble avec une attention particulière portée à leur capacité prédictive et non à la relation pouvant exister entre une variable d'exposition spécifique et l'issue.

Un total de huit scores de risque clinique a été développé, soit deux pour chacune des quatre issues. Ceux-ci ont permis d'attribuer un poids rattaché à la présence de chacun des prédicteurs qui, une fois combinés, établissent le risque pour une personne de contracter l'issue concernée. Ils sont simples à utiliser par un clinicien détenant l'information relative aux manifestations cliniques de la STB chez un patient. Le score de risque pour un individu se calcule en effectuant la somme des poids associés à la présence de chacun des prédicteurs. Le résultat obtenu peut ensuite être transposé au tableau dans lequel la probabilité de développer l'issue pour des scores individuels allant de -10 à 40 y est détaillée, un score plus élevé correspondant à un risque supérieur. L'étendue des prédictions diffère selon chaque modèle. Le modèle complet portant sur l'AML est en mesure de prédire jusqu'à un risque de 96 % de contracter l'issue, tandis que le modèle de LAM avec sélection par LASSO ne peut prédire qu'un risque maximal de 27 %.

L'une des caractéristiques principales témoignant d'une bonne capacité prédictive d'un modèle de prédiction à visée pronostique est la mesure de son habileté à discriminer entre ceux qui développeront l'issue et ceux qui ne la développeront pas. La discrimination des modèles a été estimée dans le cadre du présent projet de recherche à l'aide du C-index. Pour ce qui est des quatre modèles complets comprenant chacun des prédicteurs potentiels, la discrimination était bonne pour les trois issues concernant les AML et forte pour la prédiction de LAM. La capacité discriminative des modèles a été évaluée dans un premier temps dans les mêmes données ayant servi à leur développement et dans un second temps avec un processus de validation interne par rééchantillonnage par bootstrap. La diminution de la performance suivant la validation interne était presque inexistante pour chacun des quatre modèles. Cette faible diminution au niveau de la discrimination suggère que le score de risque clinique devrait performer de façon sensiblement



équivalente lorsque confronté à des personnes ne faisant pas partie de la TSC-NHD, à condition que l'information concernant tous les prédicteurs sélectionnés soit disponible.

Une autre caractéristique à prendre en considération lors de l'évaluation d'un modèle de prédiction est la calibration. Elle a pour nos modèles été mesurée à l'aide de courbes de calibration comparant la proportion d'issues observées dans les données à celle prédite par les modèles. La calibration s'est avérée excellente pour chacun des trois modèles complets concernant les AML. Cependant, en ce qui concerne le modèle prédisant la LAM, la calibration telle qu'estimée par la courbe s'est révélée plus modeste, bien qu'elle demeure tout de même globalement satisfaisante. À noter qu'il n'existe pas de critères clairs liés à l'évaluation qualitative de la calibration et que celle-ci dépend d'une appréciation subjective de l'allure de la courbe. Plus celle-ci se rapproche d'une pente de 1, meilleure est la calibration.

Le passage de la régression logistique à un score de risque clinique s'est fait en arrondissant les coefficients des prédicteurs. Cet arrondissement peut parfois contribuer à une diminution de la précision des estimations pouvant être faites dans les modèles. Les valeurs des C-index apparents ont été comparées entre les modèles de régression initiaux et le score de risque clinique leur étant associé afin de quantifier la baisse de capacité prédictive. Celle-ci s'est avérée minime, occasionnant des pertes ne dépassant pas un point du C-index. Cela suggère que les scores de risque clinique sont en mesure de performer de façon équivalente aux modèles de régression ayant servi à les développer.

L'application de la sélection des prédicteurs par la méthode de LASSO a permis de développer des modèles simplifiés contenant moins de prédicteurs. L'impact au niveau de la performance prédictive était négligeable, indiquant que les variables retirées n'apportaient qu'une faible contribution à la capacité prédictive des modèles. Cependant, un effet s'est fait sentir quant

à l'étendue des risques pouvant être prédits par les modèles simplifiés comparativement aux modèles complets. Le risque maximal de développement de l'issue pouvant être prédit a diminué pour chacune des quatre issues. La réduction était minimale dans le cas de l'AML (risque maximal calculé passant de 96 % à 93 %), mais plus importante pour l'AML contenant des lésions d'au moins trois centimètres de diamètre et la LAM (risque maximal calculé passant de 72 % à 55 %, et de 54 % à 27 %, respectivement). De plus, étant donné la quantité modeste de prédicteurs inclus dans le modèle complet et la facilité d'y accéder par les cliniciens traitant des patients atteints de la STB, la pertinence pratique d'un modèle simplifié peut être remise en question. Nous privilégions donc l'utilisation des modèles complets, quoiqu'il demeure intéressant d'observer le faible impact que la sélection a eu sur les mesures de performance.

## **6.2. Forces de l'étude**

Ce projet de recherche comporte plusieurs forces. Il s'agit de la première étude à avoir évalué la capacité prédictive des manifestations cliniques de la STB pour le risque de LAM et d'AML. Le tout s'est fait à l'aide de ce qui est, à ce jour, la plus large cohorte longitudinale suivant des personnes avec la STB répertoriée dans la littérature. Le vaste échantillon d'individus présentant chacune des issues, combiné au grand nombre de variables recueillies par le TSC-NHD, nous permettait d'être en mesure de développer des modèles avec une forte puissance statistique. L'ensemble du projet s'est fait de plus selon les recommandations du *Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis* (TRIPOD). Les recommandations du TRIPOD ont guidé tant les analyses que l'écriture de l'article et du présent mémoire, nous assurant de rapporter de façon aussi adéquate que possible tous les éléments devant être retrouvés dans le développement et la validation d'un modèle de prédiction. Trois des plus importants aspects de la mesure de la performance prédictive ont d'ailleurs été évalués et confirmés

par la méthode du rééchantillonnage par bootstrap, l'une des formes de validation interne les plus robustes.

## **6.3. Considérations méthodologiques**

### **6.3.1. Validité externe**

Une des limites principales affectant la présente étude relève du recrutement des participants pour la collaboration à la TSC-NHD. Celui-ci s'est fait selon un échantillonnage non probabiliste de commodité. C'est-à-dire que les chercheurs effectuant le recrutement dans chacun des centres choisissent les participants potentiels sur des bases volontaires en fonction de certains critères tels que la facilité d'accès et la disponibilité. L'utilisation d'une méthode d'échantillonnage non probabiliste comporte de nombreux désavantages pouvant entraîner des répercussions considérables sur l'interprétation des résultats.

Les deux principaux inconvénients concernent la représentativité des participants et l'impact qui en découle sur la généralisabilité des résultats. Un échantillon de participants d'une étude est dit représentatif s'il est une bonne approximation de la population qu'il représente du point de vue des aspects qui sont étudiés. Dans un contexte de développement de modèles à visée pronostique, la généralisabilité réfère à la capacité qu'a un modèle de fournir des estimations précises lorsqu'il se voit confronté à de nouvelles données [62]. En prenant en considération l'objectif principal visé par les modèles développés dans cette étude, soit qu'ils puissent être utilisés en pratique clinique ainsi qu'en recherche, il est particulièrement important que les résultats de ceux-ci puissent être généralisés à l'ensemble des gens vivant avec la STB. S'il s'avère que l'échantillon de patients recueillis par la TSC-NHD n'est pas représentatif de la population

générale avec la STB, il est possible que les résultats des modèles ne puissent pas être généralisés, ce qui compromettrait leur utilité.

Plusieurs raisons découlant du jugement subjectif des chercheurs responsables du recrutement dans chacun des centres peuvent potentiellement avoir un impact sur la représentativité des participants. L'on a qu'à penser aux patients plus facilement accessibles, plus disponibles, ceux pour qui le suivi demanderait moins d'effort ou ceux que le chercheur considérerait comme étant plus intéressants à inclure dans la base de données ont tous plus de chance d'être sélectionnés. Il devient donc particulièrement difficile de mesurer la répercussion que cela peut avoir sur les estimations faites dans les analyses. L'incapacité de déterminer l'étendue de l'impact des biais est d'ailleurs un important désavantage découlant d'une méthode d'échantillonnage non probabiliste. La réplication des résultats dans un second échantillon demeure le meilleur moyen d'en vérifier la représentativité.

## **6.3.2. Validité interne**

### **6.3.2.1. Biais de sélection**

Les biais de sélection sont ceux ayant la plus grande probabilité d'avoir eu un impact sur la présente étude, avec comme principal sous-groupe le biais d'échantillonnage, parfois nommé biais de constatation (*ascertainment bias*). Ce type de biais, qui peut survenir pour plusieurs raisons, est observé lorsque certains membres de la population visée ont une plus grande chance d'être inclus dans l'échantillon de l'étude que d'autres [48]. L'échantillon retenu est alors biaisé puisqu'il diffère systématiquement de la population qu'on souhaite étudier.

Le biais de non-couverture ou de sous-couverture, qui appartient au sous-groupe des biais de constatation, est présent dans tous les types d'étude reposant sur un échantillonnage de

commodité. Il apparaît lorsque certains sous-groupes de la population à l'étude sont exclus et ne sont pas représentés dans l'étude, ce qui rend l'échantillon recueilli non représentatif. Pour occasionner un biais, les participants doivent ne pas être représentatifs de la population source à l'égard à la fois des expositions et de l'issue. Plusieurs critères peuvent avoir une incidence sur les chances d'être invités à contribuer à la TSC-NHD. Un participant potentiel doit être suivi dans l'une des 18 cliniques participantes afin de pouvoir être invité à rejoindre la base de données. On peut dès lors raisonnablement supposer que les participants recevant un suivi dans l'une de ces cliniques diffèrent de ceux n'y étant pas suivis tant dans les manifestations cliniques de la maladie que dans les traitements et dépistages reçus. À cet effet, certaines manifestations cliniques sont plus apparentes et ont plus de chance d'être automatiquement associées à la STB, ce qui élève le risque à la fois d'obtenir un diagnostic de STB et par conséquent d'être suivi dans une des cliniques participantes. La proximité avec un des centres augmente aussi à coup sûr la probabilité d'y être suivi. Le fait d'être suivi dans l'une de ces cliniques accroît à son tour les chances qu'un dépistage d'AML soit fait par l'entremise d'une IRM, en raison de la plus grande disponibilité de cette méthode d'imagerie dans les grands centres médicaux. L'IRM étant plus précise que son alternative l'échographie dans la détection de tumeurs rénales, son utilisation accroît la possibilité qu'un AML soit localisé. En résumé, la participation à l'étude est influencée par la probabilité d'être suivi dans l'une des cliniques participantes, qui est à son tour influencée à la fois par les manifestations de la maladie et par la proximité des patients à l'une des cliniques. Ce suivi aura d'ailleurs un impact sur la méthode de dépistage employée, et donc indirectement sur le risque de détection de l'une des issues étudiées, ce qui risquerait de surestimer l'effet des prédicteurs du modèle y étant associé.

Le biais de Neyman, aussi connu sous les noms de biais de survie sélective et de biais de prévalence-incidence, survient lorsque des individus avec une manifestation plus sévère ou bien plus légère de la maladie étudiée sont exclus du bassin de participants potentiels de l'étude [48]. Dans le cadre de ce projet de recherche, nous ne pouvons pas ignorer le risque de présence de ce biais, qui dans ce cas-ci serait principalement provoqué par la combinaison de méthodes rétrospectives et prospectives dans la collecte des données. L'information rétrospective est ajoutée par les chercheurs responsables selon les dossiers médicaux de patients recrutés au sujet de diagnostics qu'ils ont reçus précédemment à leur inclusion. Sachant que tous les participants sont vivants au moment de leur recrutement, ceux ayant eu une manifestation plus sévère de la maladie ayant causé leur décès avant qu'ils aient eu la chance de participer au TSC-NHD ne contribuent pas aux renseignements rétrospectifs entrés dans la base de données. Il existe donc un risque que les cas plus graves de la maladie y soient sous-représentés. Cela peut biaiser les résultats puisque les deux issues étudiées font partie des principales causes de mortalité associées à la STB, donc une part de leur lien avec les prédicteurs peut être sous-estimée si les cas moins sévères sont plus représentés qu'ils ne le sont dans la population réelle.

Le biais dû aux risques concurrents (*competing risks*) pourrait lui aussi avoir eu un effet sur cette étude. Celui-ci est aperçu lorsqu'un événement survient, généralement la mort, avant qu'un participant n'ait eu l'occasion de développer l'issue étudiée [63]. En ce qui concerne notre étude, on peut l'observer si les individus pour qui la maladie est plus sévère durant l'enfance sont à plus haut risque de décès avant d'avoir pu obtenir un diagnostic d'AML ou de LAM. On sait que les AML sont en général diagnostiqués autour de l'âge de 10 ans tandis que les LAM le sont dans la plupart des cas plus tard à l'âge adulte [7]. On sait aussi que les risques de mortalité sont intrinsèquement liés à la sévérité de la maladie dès l'enfance [1]. On peut penser en particulier aux

rhabdomyomes cardiaques ainsi qu'à la présence d'une épilepsie sévère qui sont deux importantes causes de mortalité en bas âge associées à la STB. S'il s'avère que ces mêmes manifestations utilisées en tant que variables dans nos modèles sont liées à une augmentation du risque de développement d'AML et de LAM, une partie de l'association ne sera pas observée dans les données pour les patients ne s'étant pas rendus à un âge auquel une des deux conditions aurait pu être détectée. Il est légitime de présumer que la relation entre ces manifestations cliniques causant un décès avant la participation à la base de données et les issues étudiées est sous-estimée. On peut cependant remettre en question l'impact que cela aurait sur la généralisabilité des résultats et sur l'usage futur des scores de risque clinique. Puisque l'objectif de cette étude vise la prédiction et non l'étiologie, on peut supposer que leur risque de décès avant de développer l'une des deux conditions dans la base de données soit semblable à celui présent dans la population générale et donc que la capacité prédictive ne serait pas affectée.

Certaines des variables choisies pour le développement des modèles de prédiction comprenaient un nombre élevé de valeurs manquantes préalablement à l'imputation multiple des données. Le pourcentage de données étant manquantes se situait entre 0 % et 20 % selon la variable. Un biais de sélection dû aux données manquantes peut être observé si les participants pour qui les données sont complètes ne sont pas représentatifs de la population à l'étude [48]. Le risque qu'un biais ait été introduit par les valeurs manquantes dans cette étude est cependant faible en raison de la gestion de celles-ci par l'imputation multiple. Cette méthode demeure efficace à réduire le risque de biais même lorsque confrontée à un pourcentage de valeurs manquantes élevé, à condition que celles-ci ne soient pas manquantes de façon non aléatoire [64].

### **6.3.2.2. Biais d'information**

En plus du risque de biais de sélection, la méthode de dépistage utilisée pour détecter la présence d'AML nous place aussi à risque de biais d'information. Tel que mentionné auparavant, le diagnostic d'un AML repose principalement sur le passage d'une échographie ou d'une IRM. L'échographie est facile d'accès, abordable, et elle peut être faite dans la plupart des centres médicaux. Elle est cependant moins précise que l'IRM pour la localisation des tumeurs rénales, surtout lorsque celles-ci comprennent peu de masses graisseuses [8]. L'IRM en revanche est plus coûteuse et n'est en général disponible que dans les grands centres médicaux. Un patient avec la STB présentant une expression plus sévère de la maladie a de plus grandes chances d'être dirigé vers l'IRM qu'un qui en est affecté de façon moindre. À cet effet, les manifestations cliniques exploitées comme prédicteurs dans nos modèles sont celles principalement retrouvées chez les enfants. On peut donc émettre la supposition que le jugement ayant mené à la décision de faire passer ou non une IRM à un patient soit impacté par la présence de certains des prédicteurs employés dans cette étude. Advenant que ce soit le cas, cela nous place à risque d'erreur de classification différentielle puisque l'erreur de classification de l'issue (AML présente, mais pas détectée par l'échographie) serait influencée par la sévérité de la maladie en lien avec les manifestations cliniques de celle-ci utilisées dans le développement des modèles. L'impact de ce biais sur les résultats serait probablement une surestimation de l'effet des prédicteurs puisque l'issue serait mieux détectée chez les personnes exposées.

### **6.3.2.3. Biais de confusion**

Bien que l'évaluation de la présence de biais de confusion se veule centrale à la majorité de la recherche dans le domaine de l'épidémiologie, elle n'est habituellement pas prise en considération lors d'études portant sur le développement et la validation de modèles de prédiction



[35]. Les analyses qui ont été menées dans cette étude n'avaient pas pour objectif l'acquisition de nouvelles connaissances relativement à l'étiologie des AML et de la LAM chez les personnes avec la STB. Aucune supposition n'était par conséquent faite quant à un lien causal unissant les prédicteurs et les issues dans nos modèles. Puisque seule la corrélation entre les variables était mesurée, il n'était donc pas pertinent ni même envisageable de procéder à une évaluation de la confusion.

## **6.4. Contributions et implications cliniques**

La sclérose tubéreuse de Bourneville est une maladie relativement rare, mais pouvant avoir un impact substantiel sur la qualité de vie de ceux qui en sont atteints. Comme mentionné auparavant, les AML et la LAM figurent parmi les principales sources de mortalité chez les adultes atteints de la maladie [10, 16]. À cet effet, nous considérons que tout résultat permettant d'identifier plus tôt et de façon plus adéquate les gens à haut risque de souffrir de l'une de ces deux conditions entraînera des répercussions cliniques importantes au niveau de la prévention et du traitement pour les individus vivant avec la STB.

Un intérêt grandissant est accordé dans la recherche portant sur la STB au sujet de la modification du pronostic de la maladie. Un nombre croissant d'articles sont publiés dans le but de déterminer si la prise de médicaments servant en général à traiter des conditions présentes chez les personnes atteintes de la STB serait en mesure de reporter ou bien même de prévenir leur apparition si ceux-ci sont utilisés de manière préventive. L'un d'eux par exemple est l'essai clinique EPISTOP, dans lequel le médicament anticonvulsif vigabatrin était offert de façon prophylactique à de jeunes enfants n'ayant pas encore reçu de diagnostic d'épilepsie. Celui-ci s'est montré efficace pour retarder les premières crises convulsives et dans la réduction des risques

d'éventuellement développer une épilepsie résistante à la médication [61]. La prise de médicaments inhibiteurs de mTOR se veut le traitement de première intention pour tenter de contrôler ou de renverser les effets de l'AML et de la LAM. L'essai clinique EXIST-2 est arrivé à montrer l'efficacité de l'everolimus dans la diminution du volume des lésions d'AML et un second essai clinique a associé la prise de sirolimus à une baisse des symptômes de LAM et à une amélioration de la qualité de vie [12, 18]. La capacité d'identifier les personnes à haut risque de développer une issue se veut essentielle lors du développement d'essais cliniques portant sur l'évaluation de traitements prophylactiques. Cela permet au niveau du recrutement de tenir compte du risque d'éventuellement développer l'issue aux critères d'inclusion et d'ainsi assurer un risque de base équivalent entre chaque branche de l'étude. Pour les analyses, un ajustement selon le risque de base peut être fait dans le but d'en contrôler l'impact dans les résultats [65]. Les scores de risque clinique que nous avons développé dans ce projet de recherche permettront l'identification des participants potentiels étant à haut risque de contracter des AML ou LAM. Cela s'avérerait utile, advenant que des essais cliniques randomisés soient réalisés afin d'évaluer les bienfaits de la prise de médicaments inhibiteurs de mTOR de façon prophylactique sur le développement et les risques de complications des deux issues. Si celles-ci sont concluantes et réussissent à montrer les avantages des traitements prophylactiques pour l'AML et la LAM, les scores développés dans ce projet se verront d'autant plus utiles pour aider les cliniciens à cibler les patients les plus propices à en bénéficier.

Les recommandations en vigueur en matière de dépistage visant à détecter la présence ou bien à suivre la progression des tumeurs rénales et pulmonaires chez les personnes vivant avec la STB ne sont pas adaptées au risque individuel des patients et ne prennent pas en compte la grande hétérogénéité de la maladie. Ceux-ci sont donc contraints de devoir subir des imageries en série,

généralement jusqu'à la fin de leurs vies. En ce qui concerne l'AML, on conseille actuellement de passer une IRM au moins une fois tous les trois ans dès l'obtention d'un diagnostic de STB [8]. Pour ce qui est de la LAM, on recommande aux femmes adultes une tomodensitométrie à tous les cinq à sept ans jusqu'à la ménopause [8]. Les études publiées jusqu'à présent dans la littérature ne permettent pas une distinction du risque assez importante pour personnaliser les échéanciers de dépistage. Les scores de risque cliniques développés sont en mesure de mieux identifier le risque individuel de développer les deux conditions. Une connaissance de ce risque permettrait d'adapter le dépistage selon un échéancier plus personnalisé, assurant à la fois un meilleur suivi pour les individus à haut risque et allégeant le fardeau associé aux processus d'imagerie et de dépistage pour ceux chez qui le risque est plus faible.

## **6.5. Orientations futures**

Il est évident à la lumière des limites détaillées plus tôt dans ce chapitre que le principal enjeu menaçant l'interprétation des résultats est lié au risque que ceux-ci ne soient pas généralisables à la population complète d'individus vivant avec la STB. Il est donc primordial que les scores de risque clinique que nous avons développé soient validés de façon externe dans une seconde étude. Celle-ci devrait être basée sur un groupe de participants avec la STB, mais qui n'est pas rattaché à la TSC-NHD. Étant donné le grand niveau d'incertitude associé aux biais possibles dans nos analyses, il est difficile pour la majorité de ceux-ci de quantifier l'ampleur de l'effet qu'ils ont sur nos résultats. Une validation externe confirmant la validité de nos modèles serait le meilleur moyen de nous assurer que les biais potentiels détectés n'ont qu'un impact minime sur nos résultats. Outre l'évaluation des biais, la validation externe est une étape essentielle menant à l'adoption d'un modèle de prédiction pouvant être employé en pratique tant dans un contexte clinique que dans la recherche [21, 24]. Nous avons à cet effet entamé les démarches afin de

pouvoir effectuer une validation externe des scores développés dans ce projet en utilisant la cohorte TOSCA, qui est citée à quelques reprises dans les précédentes sections [7].

## **6.6. Conclusion**

Nous avons dans cette étude développé des scores de risque clinique permettant de prédire la probabilité de survenue d'angiomyolipomes rénaux ainsi que de lymphangioléiomyomatose pulmonaire chez des personnes vivant avec la sclérose tubéreuse. Les résultats obtenus suggèrent que les manifestations cliniques se produisant durant l'enfance sont en mesure de prédire de manière efficace le risque futur des deux complications. Bien que les résultats doivent être validés de façon externe dans une étude subséquente, nous considérons le travail effectué dans ce mémoire comme un important premier pas menant vers des soins plus personnalisés pour tous ceux atteints de la STB.

## Références bibliographiques

1. Henske, E.P., et al., *Tuberous sclerosis complex*. Nat Rev Dis Primers, 2016. **2**: p. 16035.
2. Lyczkowski, D.A., et al., *Intrafamilial phenotypic variability in tuberous sclerosis complex*. J Child Neurol, 2007. **22**(12): p. 1348-55.
3. Rakowski, S.K., et al., *Renal manifestations of tuberous sclerosis complex: Incidence, prognosis, and predictive factors*. Kidney Int, 2006. **70**(10): p. 1777-82.
4. Cudzilo, C.J., et al., *Lymphangiomyomatosis screening in women with tuberous sclerosis*. Chest, 2013. **144**(2): p. 578-585.
5. Rosset, C., C.B.O. Netto, and P. Ashton-Prolla, *TSC1 and TSC2 gene mutations and their implications for treatment in Tuberous Sclerosis Complex: a review*. Genet Mol Biol, 2017. **40**(1): p. 69-79.
6. Krueger, D.A., H. Northrup, and G. International Tuberous Sclerosis Complex Consensus, *Tuberous sclerosis complex surveillance and management: recommendations of the 2012 International Tuberous Sclerosis Complex Consensus Conference*. Pediatr Neurol, 2013. **49**(4): p. 255-65.
7. Kingswood, J.C., et al., *Tuberous Sclerosis registry to increase disease Awareness (TOSCA) - baseline data on 2093 patients*. Orphanet J Rare Dis, 2017. **12**(1): p. 2.
8. Northrup, H., et al., *Updated International Tuberous Sclerosis Complex Diagnostic Criteria and Surveillance and Management Recommendations*. Pediatr Neurol, 2021. **123**: p. 50-66.
9. Northrup, H., D.A. Krueger, and G. International Tuberous Sclerosis Complex Consensus, *Tuberous sclerosis complex diagnostic criteria update: recommendations of the 2012 International Tuberous Sclerosis Complex Consensus Conference*. Pediatr Neurol, 2013. **49**(4): p. 243-54.
10. Amin, S., et al., *Causes of mortality in individuals with tuberous sclerosis complex*. Dev Med Child Neurol, 2017. **59**(6): p. 612-617.
11. Franz, D.N., et al., *Efficacy and safety of everolimus for subependymal giant cell astrocytomas associated with tuberous sclerosis complex (EXIST-1): a multicentre, randomised, placebo-controlled phase 3 trial*. Lancet, 2013. **381**(9861): p. 125-32.
12. Bissler, J.J., et al., *Everolimus for angiomyolipoma associated with tuberous sclerosis complex or sporadic lymphangiomyomatosis (EXIST-2): a multicentre, randomised, double-blind, placebo-controlled trial*. Lancet, 2013. **381**(9869): p. 817-24.
13. French, J.A., et al., *Adjunctive everolimus therapy for treatment-resistant focal-onset seizures associated with tuberous sclerosis (EXIST-3): a phase 3, randomised, double-blind, placebo-controlled study*. The Lancet, 2016. **388**(10056): p. 2153-2163.
14. Luo, C., et al., *Perfect match: mTOR inhibitors and tuberous sclerosis complex*. Orphanet Journal of Rare Diseases, 2022. **17**(1): p. 106.
15. Flum, A.S., et al., *Update on the Diagnosis and Management of Renal Angiomyolipoma*. J Urol, 2016. **195**(4 Pt 1): p. 834-46.
16. Crino, P.B., K.L. Nathanson, and E.P. Henske, *The tuberous sclerosis complex*. N Engl J Med, 2006. **355**(13): p. 1345-56.

17. Harknett, E.C., et al., *Use of variability in national and regional data to estimate the prevalence of lymphangiomyomatosis*. QJM, 2011. **104**(11): p. 971-9.
18. McCormack, F.X., et al., *Efficacy and safety of sirolimus in lymphangiomyomatosis*. N Engl J Med, 2011. **364**(17): p. 1595-606.
19. Khawar, M.U., et al., *Clinical outcomes and survival following lung transplantation in patients with lymphangiomyomatosis*. J Heart Lung Transplant, 2019. **38**(9): p. 949-955.
20. Collins, G.S., et al., *Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement*. BMJ, 2015. **350**: p. g7594.
21. Moons, K.G., et al., *Prognosis and prognostic research: what, why, and how?* BMJ, 2009. **338**: p. b375.
22. van Smeden, M., et al., *Clinical prediction models: diagnosis versus prognosis*. J Clin Epidemiol, 2021. **132**: p. 142-145.
23. Royston, P., et al., *Prognosis and prognostic research: Developing a prognostic model*. BMJ, 2009. **338**: p. b604.
24. Moons, K.G., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration*. Ann Intern Med, 2015. **162**(1): p. W1-73.
25. Peduzzi, P., et al., *A simulation study of the number of events per variable in logistic regression analysis*. J Clin Epidemiol, 1996. **49**(12): p. 1373-9.
26. Peduzzi, P., et al., *Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates*. J Clin Epidemiol, 1995. **48**(12): p. 1503-10.
27. Vittinghoff, E. and C.E. McCulloch, *Relaxing the rule of ten events per variable in logistic and Cox regression*. Am J Epidemiol, 2007. **165**(6): p. 710-8.
28. Steyerberg, E.W., et al., *Internal validation of predictive models: efficiency of some procedures for logistic regression analysis*. J Clin Epidemiol, 2001. **54**(8): p. 774-81.
29. Bozdogan, H., *Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions*. Psychometrika, 1987. **52**(3): p. 345-370.
30. Steyerberg, E.W., et al., *Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets*. Med Decis Making, 2001. **21**(1): p. 45-56.
31. Tibshirani, R., *Regression Shrinkage and Selection via the Lasso*. Journal of the Royal Statistical Society. Series B (Methodological), 1996. **58**(1): p. 267-288.
32. Altman, D.G., et al., *Prognosis and prognostic research: validating a prognostic model*. BMJ, 2009. **338**: p. b605.
33. Bonnett, L.J., et al., *Guide to presenting clinical prediction models for use in clinical settings*. BMJ, 2019. **365**: p. l737.
34. Jehi, L., et al., *Development and validation of nomograms to provide individualised predictions of seizure outcomes after epilepsy surgery: a retrospective analysis*. Lancet Neurol, 2015. **14**(3): p. 283-90.
35. van Diepen, M., et al., *Prediction versus aetiology: common pitfalls and how to avoid them*. Nephrology Dialysis Transplantation, 2017. **32**(suppl\_2): p. ii1-ii5.
36. Kothare, S.V., et al., *Severity of manifestations in tuberous sclerosis complex in relation to genotype*. Epilepsia, 2014. **55**(7): p. 1025-9.

37. Muzykewicz, D.A., et al., *TSC1 and TSC2 mutations in patients with lymphangiomyomatosis and tuberous sclerosis complex*. J Med Genet, 2009. **46**(7): p. 465-8.
38. Ryu, J.H., et al., *Cystic lung disease is not uncommon in men with tuberous sclerosis complex*. Respir Med, 2012. **106**(11): p. 1586-90.
39. *Tsc Alliance - Natural History Database*. 2023; Available from: <https://www.tscalliance.org/researchers/natural-history-database/>.
40. Greenland, S., J. Pearl, and J.M. Robins, *Causal diagrams for epidemiologic research*. Epidemiology, 1999. **10**(1): p. 37-48.
41. Yamakado, K., et al., *Renal angiomyolipoma: relationships between tumor size, aneurysm formation, and rupture*. Radiology, 2002. **225**(1): p. 78-82.
42. Van Buuren, S. and K. Groothuis-Oudshoorn, *mice: Multivariate imputation by chained equations in R*. Journal of statistical software, 2011. **45**: p. 1-67.
43. Robitzsch, A. and S. Grund, *miceadds: Some Additional Multiple Imputation Functions, Especially for 'mice'*. 2023.
44. Friedman, J., T. Hastie, and R. Tibshirani, *Regularization Paths for Generalized Linear Models via Coordinate Descent*. J Stat Softw, 2010. **33**(1): p. 1-22.
45. Harrell Jr, F.E., *rms: Regression Modeling Strategies*. 2022.
46. Sterne, J.A.C., et al., *Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls*. BMJ, 2009. **338**: p. b2393.
47. Harrell, J., Frank E and F.E. Harrell, *Missing data*. Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis, 2015: p. 45-61.
48. Delgado-Rodríguez, M. and J. Llorca, *Bias*. Journal of Epidemiology and Community Health, 2004. **58**(8): p. 635-641.
49. Azur, M.J., et al., *Multiple imputation by chained equations: what is it and how does it work?* International journal of methods in psychiatric research, 2011. **20**(1): p. 40-49.
50. White, I.R., P. Royston, and A.M. Wood, *Multiple imputation using chained equations: issues and guidance for practice*. Statistics in medicine, 2011. **30**(4): p. 377-399.
51. Moons, K.G., et al., *Using the outcome for imputation of missing predictor values was preferred*. Journal of clinical epidemiology, 2006. **59**(10): p. 1092-1101.
52. Steyerberg, E.W., et al., *Prognosis Research Strategy (PROGRESS) 3: prognostic model research*. PLoS Med, 2013. **10**(2): p. e1001381.
53. Steyerberg, E.W., M.J. Eijkemans, and J.D.F. Habbema, *Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis*. Journal of clinical epidemiology, 1999. **52**(10): p. 935-942.
54. Steyerberg, E.W., et al., *Poor performance of clinical prediction models: the harm of commonly applied methods*. J Clin Epidemiol, 2018. **98**: p. 133-143.
55. Steyerberg, E.W., et al., *Internal and external validation of predictive models: a simulation study of bias and precision in small samples*. Journal of clinical epidemiology, 2003. **56**(5): p. 441-447.
56. Vergouwe, Y., et al., *Development and validation of a prediction model with missing predictor data: a practical approach*. Journal of clinical epidemiology, 2010. **63**(2): p. 205-214.

57. Steyerberg, E.W., *Overfitting and optimism in prediction models*, in *Clinical prediction models*. 2019, Springer. p. 95-112.
58. Steyerberg, E.W., et al., *Assessing the performance of prediction models: a framework for some traditional and novel measures*. *Epidemiology (Cambridge, Mass.)*, 2010. **21**(1): p. 128.
59. Hosmer Jr, D.W., S. Lemeshow, and R.X. Sturdivant, *Applied logistic regression*. Vol. 398. 2013: John Wiley & Sons.
60. van Buuren, S. and K. Groothuis-Oudshoorn, *mice: Multivariate Imputation by Chained Equations in R*. *Journal of Statistical Software*, 2011. **45**(3): p. 1 - 67.
61. Kotulska, K., et al., *Prevention of Epilepsy in Infants with Tuberous Sclerosis Complex in the EPISTOP Trial*. *Ann Neurol*, 2021. **89**(2): p. 304-314.
62. Justice, A.C., K.E. Covinsky, and J.A. Berlin, *Assessing the generalizability of prognostic information*. *Ann Intern Med*, 1999. **130**(6): p. 515-24.
63. Wolbers, M., et al., *Prognostic Models With Competing Risks: Methods and Application to Coronary Risk Prediction*. *Epidemiology*, 2009. **20**(4): p. 555-561.
64. Van Buuren, S., *Flexible imputation of missing data*. 2018: CRC press.
65. Steyerberg, E.W. and E.W. Steyerberg, *Applications of prediction models*. *Clinical prediction models: a practical approach to development, validation, and updating*, 2019: p. 15-36.



# **Annexes**

## **Annexe 1. Mini-protocole**

Université de Montréal

Mini-protocole

Prédiction du développement des angiomyolipomes rénaux et de la lymphangioléiomyomatose pulmonaire chez les gens atteints de la sclérose tubéreuse de Bourneville

Par :

Frédéric Loubert

Directeur de recherche :

Mark R. Keezer

École de santé publique de l'Université de Montréal (ESPUM)

28 août 2022

**Contexte, revue de la littérature et justification.** La sclérose tubéreuse de Bourneville (STB) est une maladie génétique qui affecte approximativement une naissance vivante sur 6000<sup>1</sup>. Elle est une des causes génétiques primaires de l'épilepsie. Chez l'enfant, la principale source de morbidité provient de manifestations survenant au niveau du cœur, du cerveau ainsi que de la peau. Quant aux adolescents et aux adultes, ils sont à plus haut risque de développer des angiomyolipomes rénaux (AML) et de la lymphangioliomyomatose pulmonaire (LAM)<sup>1</sup>.

La STB est reconnue en partie pour son importante hétérogénéité phénotypique, que l'on peut retrouver chez les membres d'une même famille présentant la même mutation génétique<sup>2</sup>. Le fardeau associé à la maladie est aussi largement affecté par l'âge de l'individu. Il n'existe présentement que très peu de moyens d'aider les médecins à déterminer lesquels de leurs patients atteints de la STB sont les plus à risque de développer l'AML ou la LAM. Les études publiées à ce jour se sont limitées à montrer que l'âge, le sexe féminin, et les mutations génétiques du gène TSC2 comparativement au gène TSC1 sont associés à un risque plus élevé d'AML et de LAM.<sup>3,4</sup>

<sup>5</sup> Les recommandations en matière de dépistage pour les patients atteints de la STB traitent chaque individu de manière égale et ignorent complètement la grande hétérogénéité de la maladie. Les patients sont donc soumis à des imageries en série, et ce jusqu'à la fin de leurs vies.<sup>6</sup> Il n'existe présentement pas de données dans la littérature visant à déterminer s'il existe des caractéristiques plus discriminatives basées sur des manifestations cliniques de la maladie à l'enfance nous permettant d'identifier les adolescents ou les adultes à haut risque de développer l'une des deux complications.

Une meilleure capacité à prédire quels adultes atteints de STB sont à plus haut risque de développer l'AML et la LAM permettrait la création de protocoles de dépistage personnalisés et

adaptés au risque individuel de chacun et de cerner les individus pouvant bénéficier de nouvelles thérapies préventives telles que la prise d'inhibiteurs de mTOR.

**Objectifs.** Ce projet de recherche a pour principal objectif de déterminer la capacité du fardeau de la maladie au niveau cutané et au niveau du cerveau chez l'enfant de prédire le développement futur ainsi que la sévérité de l'AML rénal et de la LAM pulmonaire chez l'adolescent et chez l'adulte.

**Devis d'étude.** Ceci est une étude de développement et de validation interne de modèles de prédiction utilisant une cohorte longitudinale avec collecte de données ambispective (issue d'une combinaison de méthodes à la fois prospectives et rétrospectives).

**Gestion des données et stratégie d'analyse.** Les données nécessaires à la réalisation de ce projet proviennent de la TSC Alliance Natural History Database (TSC-NHD). La TSC-NHD contient de l'information sur 2420 individus consentants atteints de la STB portant sur les aspects sociodémographiques, les manifestations cliniques de la maladie, les précédents résultats d'imagerie et les traitements reçus ainsi que leur efficacité.

Les données utilisées pour ce projet contiendront quatre issues : l'AML, l'AML bilatéral, l'AML avec lésions de plus de 3 cm et la LAM. D'autre part, les variables avec lesquelles ces issues seront prédites incluront les principales manifestations cliniques de la STB au niveau du cœur, du cerveau, de la peau et des yeux ainsi que l'âge, le sexe et la mutation génétique (liste détaillée au Tableau 1).

Le modèle de risques instantanés proportionnels de Cox sera utilisé afin de développer les modèles servant à prédire l'association entre les prédicteurs et les issues d'intérêt. Son utilisation permettra d'accommoder pour la durée de suivi qui varie selon les participants. Pour chaque issue étudiée, un second modèle sera envisagé dans lequel LASSO sera utilisé pour la sélection de

variables. Avec chaque modèle final, un poids sera attribué à chacun des prédicteurs. Celui-ci sera obtenu en divisant individuellement chaque coefficient du modèle initial par la plus petite valeur obtenue de tous les coefficients. La somme des poids pour un individu représentera son score de prédiction.

L'imputation multiple par équations chaînées (Mice) sera privilégiée afin de gérer les valeurs manquantes dans les données. Dix bases de données distinctes seront ainsi créées et des analyses combinées seront menées sur les résultats des imputations.

La validation interne des modèles de prédiction sera assurée à l'aide de la méthode de rééchantillonnage par Bootstrap. Celle-ci nous permettra de ne pas avoir à diviser l'échantillon en deux et ainsi de pouvoir bénéficier de la pleine cohorte. Cette méthode est validée tant pour les modèles de régression logistique<sup>7,8</sup> que pour les modèles de régression de Cox<sup>9,10</sup>.

La performance prédictive des deux modèles (complet versus parcimonieux) pour chaque issue sera mesurée en utilisant la discrimination et la calibration. La discrimination correspond à la capacité qu'a un modèle de distinguer entre les individus qui développeront l'issue et ceux qui ne la développeront pas et sera menée en utilisant le C-index de Harrell. La calibration correspond à l'accord entre l'issue observée et l'issue prédite par le modèle et est illustrée à l'aide d'une courbe de calibration. Ces méthodes de mesure de la validité prédictive sont toutes deux applicables à la régression de Cox<sup>11</sup>.

**Taille d'échantillon.** Des simulations ont précédemment montré que de 20 à 40 événements par variable, lorsque le rééchantillonnage par Bootstrap est utilisé pour la validation interne d'un modèle de régression logistique, suffisent à obtenir des estimations stables avec un niveau de biais faible<sup>12</sup>. Bien que de telles simulations n'aient pas été répétées pour la régression de Cox, il a été

montré que le nombre d'événements par variable requis pour la régression de Cox est habituellement le même que celui requis pour les régressions logistiques.

Les données du TSC-NHD comprennent 2420 participants, desquels 426 ont développé un AML après l'âge de 10 ans, portant à 33 le nombre d'événements pour chacun des 13 prédicteurs. Pour la LAM, 77 participants l'ont développé, portant à 6 le nombre d'événements par variable.

**Biais potentiels.** Une source potentielle de biais d'information réside dans la manière dont est diagnostiqué l'AML. Le diagnostic peut se faire à l'aide d'une IRM ou d'une échographie, le premier étant plus précis mais aussi plus coûteux, le second plus accessible mais moins précis. Cela représente un risque d'erreur de classification différentielle si le choix de l'IRM se voit associé à la sévérité de la maladie en lien avec d'autres manifestations étudiées dans ce projet.

Certaines des manifestations cliniques utilisées en tant que prédicteurs dans l'étude augmentent le risque qu'un individu reçoive un diagnostic de STB et qu'il soit suivi dans une clinique participant au TSC-NHD. C'est le cas de certaines manifestations cutanées et des lésions au cerveau. De plus, le statut socioéconomique influence la probabilité qu'un participant réside dans une ville à proximité de grands centres médicaux, augmentant à la fois la probabilité de recevoir une IRM et d'être recruté pour participer au TSC-NHD. La participation à la banque de données est donc possiblement un collisionneur puisqu'elle est influencée à la fois par certains des prédicteurs et par l'issue.

Certains marqueurs génétiques étant des causes communes à la fois de l'AML et de certaines manifestations cutanées ne sont pas prises en compte dans les données de la TSC-NHD, ce qui pourrait potentiellement causer de la confusion. La présence de mutation génétique s'arrêtant à la présence ou bien à l'absence des gènes *TSC1* et *TSC2* est plutôt large et nous place à risque de confusion résiduelle.

Une nuance existe dans l'interprétation du risque de biais, et plus particulièrement de la confusion, lorsqu'appliqué aux études de développement de modèles de prédiction comparativement aux études à visées étiologiques.<sup>13</sup> À cet effet, l'outil PROBAST sera envisagé afin d'évaluer la présence de biais dans notre étude.<sup>14</sup>

**Considérations éthiques.** Toutes les données incluses dans le TSC-NHD sont dénominalisées. Il n'est donc pas possible d'identifier les participants dans la base de données. Seuls les membres de l'équipe de recherche auront accès aux données, et ce seulement à des fins de recherche. Seuls les résultats finaux des analyses seront publiés dans un journal révisé par les pairs.

**Faisabilité.** La demande d'accès aux données est déjà approuvée et toutes les données nécessaires à la réalisation du projet nous ont été transmises. De plus, le projet de recherche a été approuvé par le CER du CHUM. Le nettoyage des données et la définition des variables sont en cours.

## Calendrier 2022-2023

- **Septembre :**
  - Préparation de la présentation du protocole pour le 8 septembre.
  - Poursuite du nettoyage des données et de la définition des variables.
  - Finalisation du protocole complet.
- **Octobre :**
  - Analyses descriptives des données.
  - Début de la rédaction du mémoire (introduction et recension des écrits).
- **Novembre :**
  - Analyses statistiques principales des données.
  - Rédaction des sections des objectifs et des méthodes du mémoire.
  - Révision des sections complétées du mémoire.
- **Décembre :**
  - Finalisation des analyses sur les données.
  - Poursuite de la rédaction des méthodes.
  - Révision des sections précédemment complétées.
- **Janvier :**
  - Rédaction de la section des résultats.
  - Révision des sections précédemment complétées.
  - Rédaction de l'article scientifique.
- **Février :**
  - Rédaction de la section de discussion et de la conclusion du mémoire.
  - Révision des sections précédemment complétées.

- Soumission de l'article pour publication.
- **Mars :**
  - Finalisation de la rédaction.
  - Révision de toutes les sections du mémoire.
  - Remise de l'avis de dépôt du mémoire.
- **Avril :**
  - Révisions finales du mémoire en entier.
  - Mise en page finale précédant la soumission.
  - Présentation de l'affiche scientifique au congrès annuel de L'American Academy of Neurology.
- **Mai :**
  - Soumission du mémoire.
- **Juin :**
  - Corrections du mémoire s'il y a lieu.



## **Références**

1. Henske, E. P., Jozwiak, S., Kingswood, J. C., Sampson, J. R., & Thiele, E. A. (2016). Tuberous sclerosis complex. *Nat Rev Dis Primers*, 2, 16035. <https://doi.org/10.1038/nrdp.2016.35>
2. Lyczkowski, D. A., Conant, K. D., Pulsifer, M. B., Jarrett, D. Y., Grant, P. E., Kwiatkowski, D. J., & Thiele, E. A. (2007). Intrafamilial phenotypic variability in tuberous sclerosis complex. *J Child Neurol*, 22(12), 1348-1355. <https://doi.org/10.1177/0883073807307093>
3. Rakowski, S. K., Winterkorn, E. B., Paul, E., Steele, D. J., Halpern, E. F., & Thiele, E. A. (2006). Renal manifestations of tuberous sclerosis complex: Incidence, prognosis, and predictive factors. *Kidney Int*, 70(10), 1777-1782. <https://doi.org/10.1038/sj.ki.5001853>
4. Cudzilo, C. J., Szczesniak, R. D., Brody, A. S., Rattan, M. S., Krueger, D. A., Bissler, J. J., Franz, D. N., McCormack, F. X., & Young, L. R. (2013). Lymphangiomyomatosis screening in women with tuberous sclerosis. *Chest*, 144(2), 578-585. <https://doi.org/10.1378/chest.12-2813>
5. Rosset, C., Netto, C. B. O., & Ashton-Prolla, P. (2017). TSC1 and TSC2 gene mutations and their implications for treatment in Tuberous Sclerosis Complex: a review. *Genet Mol Biol*, 40(1), 69-79. <https://doi.org/10.1590/1678-4685-GMB-2015-0321>
6. Krueger, D. A., Northrup, H., & International Tuberous Sclerosis Complex Consensus, G. (2013). Tuberous sclerosis complex surveillance and management: recommendations of the 2012 International Tuberous Sclerosis Complex Consensus Conference. *Pediatr Neurol*, 49(4), 255-265. <https://doi.org/10.1016/j.pediatrneurol.2013.08.002>
7. Altman, D. G., Vergouwe, Y., Royston, P., & Moons, K. G. (2009). Prognosis and prognostic research: validating a prognostic model. *BMJ*, 338, b605. <https://doi.org/10.1136/bmj.b605>
8. Steyerberg, E. W., Harrell, F. E., Jr., Borsboom, G. J., Eijkemans, M. J., Vergouwe, Y., & Habbema, J. D. (2001). Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*, 54(8), 774-781. [https://doi.org/10.1016/s0895-4356\(01\)00341-9](https://doi.org/10.1016/s0895-4356(01)00341-9)
9. Sauerbrei, W., & Schumacher, M. (1992). A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat Med*, 11(16), 2093-2109. <https://doi.org/10.1002/sim.4780111607>
10. Mojtahedi, M., Newton, S., & Von Meding, J. (2017). Predicting the resilience of transport infrastructure to a natural disaster using Cox's proportional hazards regression model. *Natural Hazards*, 85(2), 1119-1133.
11. Royston, P., & Altman, D. G. (2013). External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol*, 13, 33. <https://doi.org/10.1186/1471-2288-13-33>
12. Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol*, 165(6), 710-718. <https://doi.org/10.1093/aje/kwk052>
13. van Diepen, M., Ramspek, C. L., Jager, K. J., Zoccali, C., & Dekker, F. W. (2017). Prediction versus aetiology: common pitfalls and how to avoid them. *Nephrology Dialysis Transplantation*, 32(suppl\_2), ii1-ii5.
14. Wolff, R. F., Moons, K., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., Mallett, S., & PROBAST Group† (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of internal medicine*, 170(1), 51–58. <https://doi.org/10.7326/M18-1376>

## Annexes

**Tableau 1: Treize prédicteurs potentiels.**

<b>Variable</b>	<b>Opérationnalisation *</b>
Âge	Années
Sexe	Hommes vs. Femmes
Mutation génétique	TSC1 vs. TSC2 vs. Inconnue/non mesurée
Déficience intellectuelle	Sévère vs. Modérée vs. Aucune vs. Inconnue/non mesurée
Épilepsie (Incluant les spasmes épileptiques)	Présence vs. Absence
Tubers corticaux	Présence vs. Absence
Tumeurs sous-épendymaires à cellules géantes (SEGA)	Présence vs. Absence
Nodules sous-épendymaires	Présence vs. Absence
Angiofibromes faciaux	Présence vs. Absence
Macules hypopigmentées	≥ 3 vs. < 3
Peau de chagrin	Présence vs. Absence
Rhabdomyomes cardiaques	Présence vs. Absence
Hamartomes réiniens	Présence vs. Absence

\* Selon les variables telles que collectées par le TSC-NHD

## **Annexe 2. Commentaires de Simon De Montigny suivant la présentation du mini-protocole**

Commentaires suite à la présentation de mini-protocole

- Simon de Montigny

1. Il faut bien expliquer la nature des données ambispectives au niveau de la base de données utilisée, et distinguer cela des données qui seront concrètement utilisées pour la modélisation.
2. Il faut bien expliquer pourquoi un modèle parsimonieux serait utile, en référant à la quantité de données ( $n$ ) et le nombre de prédicteurs envisagés ( $p$ ).
3. Il faut distinguer la classification (faire un événement ou pas) du principe de survie (dans un modèle de base, tous les sujets feront l'événement éventuellement). Il peut être intéressant de lire au sujet des modèles de Cox avec « cure rate » ou avec une proportion a priori de « non-susceptibles ».

# Annexe 3. Commentaires de Nathalie Auger suivant la présentation du mini-protocole

Nom de l'étudiant: Loubert

	Exceptionnel – 5	Excellent – 4	Compétent – 3	Nécessite améliorations – 2	Incomplet - 1	Note sur 5
<b>Contexte, revue de la littérature et justification</b>	Le contexte et la revue de la littérature sont pertinents, informatifs et présentée de façon logique, incluant une critique des recherches antérieures et une identification des lacunes justifiant l'objectif proposé.	Le contexte et/ou la revue de la littérature sont assez pertinents; les lacunes dans les connaissances et une critique des recherches antérieures sont présentés, mais ils ne sont pas totalement liés aux objectifs proposés.	Le contexte et/ou la revue de la littérature sont moins clairs; les lacunes dans les connaissances doivent être déduit ou une critique des recherches antérieures justifiant l'objectif manque.	Le contexte et/ou la revue de la littérature ne sont pas clairs et ne permettent pas de situer l'objectif de la recherche dans leur contexte.	La revue de la littérature ne fournit pas les informations pertinentes.	
<b>Objectif</b>	L'objectif est clair et concis, informe le lecteur du but exact de l'étude, utilise l'approche PICOT, et n'inclut pas les méthodes.	L'objectif est assez clair et informe le lecteur du but exact de l'étude, mais omet certaines informations nécessaires (ex., la population, l'exposition, l'issue).	L'objectif omet plusieurs informations nécessaires (ex., la population, l'exposition, l'issue) mais le but de l'étude peut être déduit.	L'objectif n'est pas clair et n'inclut pas les éléments permettant de comprendre le but de l'étude.	L'objectif est manquant.	
<b>Devis d'étude</b>	Le devis du projet est approprié et clairement énoncé et décrit; les participants inclus sont bien spécifiés.	Le devis est approprié, bien énoncé mais sa description manque de clarté ou les participants ne sont pas clairement spécifiés ou décrits.	Le devis est approprié, bien énoncé mais sa description manque de clarté et les participants ne sont pas clairement spécifiés ou décrits.	Le devis est énoncé mais n'est pas du tout décrit et il manque de précisions importantes sur les participants à l'étude.	Le devis est inapproprié ou manquant.	
<b>Méthodes</b>	La description des méthodes est bien organisée et concise tout en fournissant suffisamment d'information pour permettre la reproduction des analyses; les variables primaires sont clairement définies et l'analyse statistique proposée est appropriée; les considérations de tailles d'échantillons sont bien justifiées et étayées.	Idem au niveau 5, mais contient quelques informations inutiles et/ou des descriptions moins claires.	Présente des méthodes reproductibles, avec des informations pertinentes, mais certaines informations sont manquantes et/ou les informations sont désorganisées.	Présente des méthodes qui sont légèrement reproductibles; certaines parties des méthodes doivent être déduites par le lecteur; les variables et les considérations de taille d'échantillon ne sont pas bien décrites.	Les méthodes ne sont pas cohérentes avec le devis ou sont décrits de manière qu'elles ne peuvent pas être reproduites; les variables et les considérations de taille d'échantillon sont inadéquatement décrites.	
<b>Considérations de biais potentiels</b>	Les principales sources de biais (sélection, information, confusion) sont évaluées; une proposition est fournie sur la façon dont l'étude aborde chacune d'entre elles.	Les principales sources de biais sont reconnues; une proposition est fournie sur la manière dont l'étude abordera la plupart d'entre elles.	Certaines sources de biais majeurs sont reconnues mais au moins une limite principale est manquante; une proposition est fournie sur la manière dont l'étude abordera la plupart d'entre elles.	Plusieurs sources de biais majeurs sont omises et/ou une réponse vague est fournie sur la façon dont l'étude les abordera.	Le potentiel pour le biais n'est pas adressé.	

**Commentaires sur le titre, les considérations éthiques, la faisabilité, et le calendrier proposé**

This is an excellent/exceptional protocol. The background leading into the objective is clear, good objective, methods very clear. No problems with ethics, feasibility and timeline. Overall a great protocol, this student is on the right path. Oral presentation: Objective slide not clear - write down a clear objective. eg. The objective of this study is to xxx. Assumptions of Cox regression should be explicitly stated, including variables needed for analysis. eg, describe time-to-event, competing event (eg death), censoring. Excellent presentation.

**Autres commentaires pertinents**

Minor issues: Had a hard time following the study power section - method used not clear. Also, normally we state some thing like: This study has 80% power to detect a hazard ratio of xxx. Spell out IRM at first use. Bias second paragraph - address selection bias directly. What is Probst. Table 1 - definition of deficiency intellectuelle unclear.

## Annexe 4. Certificat d'éthique émis par le Comité d'éthique de la recherche du CHUM



Comité d'éthique de la recherche du CHUM  
Pavillon R, 900 rue St-Denis, 3<sup>e</sup> étage  
Montréal (Québec) H2X 0A9

Le 27 novembre 2021

Docteur Mark Keezer

a/s: M. Frédéric Loubert  
courriel: frederic.loubert@umontreal.ca

<b>Objet:</b>	<b>21.289 – Approbation initiale-FINALE (évaluation déléguée)</b>
	<b>Prédire le développement des angiomyolipomes rénaux et des lymphangioliéiomyomatoses pulmonaires chez les gens atteints de la sclérose tubéreuse de Bourneville</b>

Docteur,

Nous accusons réception, des documents soumis en vue de l'approbation du projet mentionné en rubrique. Votre projet a été évalué en évaluation déléguée compte tenu qu'il s'agit d'un projet considéré à risque minimal pour les participants. Nous avons examiné les documents suivants:

- Formulaire de demande d'évaluation éthique d'un projet (formulaire 11)
- Formulaire Évaluation Budgétaire et Budget [date : 22 octobre 2021]
- Formulaire Évaluation Budgétaire et Budget (Budget projet) [date : 22 octobre 2021]
- Autres documents (Consentement TSC Alliance) [version : 1]
- Autres documents (Protocole TSC-NH) [version : 1]
- Curriculum Vitae (Mark Keezer)
- Autres documents (Support letter TSC Alliance) [version : 1]
- Protocole de recherche [date : 21 novembre 2021, version 2]
- Soutien Service/Département du CHUM

Le CÉR du CHUM confirme également que, possédant l'expertise scientifique nécessaire, il a également procédé à l'évaluation de la qualité et de la pertinence scientifique de projet, conformément au mandat qu'il a reçu de l'établissement qui l'a constitué.

Le tout étant jugé satisfaisant, il nous fait plaisir de vous informer que la présente constitue l'approbation finale de votre projet de recherche, **valide pour un an à compter du 27 novembre 2021.**