



**HAL**  
open science

# Algo CEPPPIA – Prédiction du risque de développement de maladies chroniques : validation d'un algorithme par corrélation avec un comité d'experts

Romain Mortier

## ► To cite this version:

Romain Mortier. Algo CEPPPIA – Prédiction du risque de développement de maladies chroniques : validation d'un algorithme par corrélation avec un comité d'experts. Médecine humaine et pathologie. 2021. dumas-03614895

**HAL Id: dumas-03614895**

<https://dumas.ccsd.cnrs.fr/dumas-03614895v1>

Submitted on 21 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ CLERMONT AUVERGNE  
UFR DE MÉDECINE ET DES PROFESSIONS PARAMÉDICALES

THÈSE D'EXERCICE

pour le

DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE

par

MORTIER Romain Thomas

Présentée et soutenue publiquement le 27/10/2021

Algo CEPPPIA – Prédiction du risque de développement de maladies chroniques :  
validation d'un algorithme par corrélation avec un comité d'experts

Directeur de thèse : Monsieur RICHARD Ruddy, Professeur, CHU Clermont-Ferrand,  
Centre de Recherche en Nutrition Humaine

Président du jury : Monsieur GERBAUD Laurent, Professeur, Pôle de Santé Publique du  
CHU Clermont-Ferrand et Service de Santé Universitaire de Clermont-Ferrand

Membres du jury :

- Monsieur DAUPLAT Jacques, Professeur honoraire, Centre Jean PERRIN
- Monsieur PEREIRA Bruno, Docteur en Biostatistiques, CHU Clermont-Ferrand,  
Direction de la Recherche Clinique et de l'Innovation





UNIVERSITÉ CLERMONT AUVERGNE  
UFR DE MÉDECINE ET DES PROFESSIONS PARAMÉDICALES

THÈSE D'EXERCICE

pour le

DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE

par

MORTIER Romain Thomas

Présentée et soutenue publiquement le 27/10/2021

Algo CEPPPIA – Prédiction du risque de développement de maladies chroniques :  
validation d'un algorithme par corrélation avec un comité d'experts

Directeur de thèse : Monsieur RICHARD Ruddy, Professeur, CHU Clermont-Ferrand,  
Centre de Recherche en Nutrition Humaine

Président du jury : Monsieur GERBAUD Laurent, Professeur, Pôle de Santé Publique du  
CHU Clermont-Ferrand et Service de Santé Universitaire de Clermont-Ferrand

Membres du jury :

- Monsieur DAUPLAT Jacques, Professeur honoraire, Centre Jean PERRIN
- Monsieur PEREIRA Bruno, Docteur en Biostatistiques, CHU Clermont-Ferrand,  
Direction de la Recherche Clinique et de l'Innovation



## UNIVERSITE CLERMONT AUVERGNE

---

PRESIDENTS HONORAIRES	: <b>JOYON</b> Louis
UNIVERSITE D'AUVERGNE	: <b>DOLY</b> Michel
	: <b>TURPIN</b> Dominique
	: <b>VEYRE</b> Annie
	: <b>DULBECCO</b> Philippe
	: <b>ESCHALIER</b> Alain
PRESIDENTS HONORAIRES	: <b>CABANES</b> Pierre
UNIVERSITE BLAISE PASCAL	: <b>FONTAINE</b> Jacques
	: <b>BOUTIN</b> Christian
	: <b>MONTEIL</b> Jean-Marc
	: <b>ODOUARD</b> Albert
	: <b>LAVIGNOTTE</b> Nadine
PRESIDENT DE L'UNIVERSITE et	
PRESIDENT DU CONSEIL ACADEMIQUE PLENIER	: <b>BERNARD</b> Mathias
PRESIDENT DU CONSEIL ACADEMIQUE RESTREINT	: <b>DEQUIEDT</b> Vianney
VICE-PRESIDENTE DU CONSEIL D'ADMINISTRATION	: <b>FOGLI</b> Anne
VICE-PRESIDENT DE LA COMMISSION DE LA RECHERCHE	: <b>HENRARD</b> Pierre
VICE PRESIDENTE DE LA COMMISSION DE LA	
FORMATION ET DE LA VIE UNIVERSITAIRE	: <b>PEYRARD</b> Françoise
DIRECTEUR GENERAL DES SERVICES	: <b>PAQUIS</b> François



## UFR DE MEDECINE ET DES PROFESSIONS PARAMEDICALES

DOYENS HONORAIRES	: <b>DETEIX</b> Patrice
	: <b>CHAZAL</b> Jean
DOYEN	: <b>CLAVELOU</b> Pierre
RESPONSABLE ADMINISTRATIVE	: <b>ROBERT</b> Gaëlle

## **LISTE DU PERSONNEL ENSEIGNANT**

### **PROFESSEURS HONORAIRES :**

MM. BACIN Franck - BEGUE René-Jean - BEYTOUT Jean - BOMMELAER Gilles - BOUCHER Daniel - BUSSIERE Jean-Louis - CANO Noël - CASSAGNES Jean - CATILINA Pierre - CHABANNES Jacques - CHAZAL Jean - CHIPPONI Jacques - CHOLLET Philippe - CITRON Bernard - COUDERT Jean - DASTUGUE Bernard - DAUPLAT Jacques - DECHELOTTE Pierre - DEMEOCQ François - DE RIBEROLLES Charles - DETEIX Patrice - ESCANDE Georges - Mme FONCK Yvette - MM. GENTOU Claude - Mme GLANDDIER Phyllis - MM. IRTHUM Bernard - JACQUETIN Bernard - Mme LAVARENNE Jeanine - MM. LAVERAN Henri - LESOURD Bruno - LEVAI Jean-Paul - MAGE Gérard - MARCHEIX Jean- Claude - MICHEL Jean-Luc - MONDIE Jean-Michel - PHILIPPE Pierre - PLAGNE Robert - PLANCHE Roger - PONSONNAILLE Jean - REY Michel - Mme RIGAL Danièle - MM. ROZAN Raymond - SCHOEFFLER Pierre - SIROT Jacques - RIBAL Jean-Pierre - SOUTEYRAND Pierre - TANGUY Alain - TERVER Sylvain - THIEBLOT Philippe - TOURNILHAC Michel - VANNEUVILLE Guy - VIALLET Jean-François - Mlle VEYRE Annie

### **PROFESSEURS EMERITES :**

MM. AUMAITRE Olivier - BOITEUX Jean-Paul - CHAMOUX Alain - DUBRAY Claude - ESCHALIER Alain - KEMENY Jean-Louis - LABBE André - Mme LAFEUILLE Hélène - MM. LEMERY Didier - LUSSON Jean-René

## **PROFESSEURS DES UNIVERSITES-PRATICIENS HOSPITALIERS**

### **PROFESSEURS DE CLASSE EXCEPTIONNELLE**

M. VAGO Philippe	Histologie-Embryologie Cytogénétique
M. AVAN Paul	Biophysique et Traitement de l'Image
M. DURIF Franck	Neurologie
M. BOIRE Jean-Yves	Biostatistiques, Informatique Médicale et Technologies de Communication
M. BOYER Louis	Radiologie et Imagerie Médicale option Clinique
M. POULY Jean-Luc	Gynécologie et Obstétrique
M. CANIS Michel	Gynécologie-Obstétrique
Mme PENAULT-LLORCA Frédérique	Anatomie et Cytologie Pathologiques
M. BAZIN Jean-Etienne	Anesthésiologie et Réanimation Chirurgicale
M. BIGNON Yves Jean	Cancérologie option Biologique
M. BOIRIE Yves	Nutrition Humaine
M. CLAVELOU Pierre	Neurologie
M. GILAIN Laurent	O.R.L.
M. LEMAIRE Jean-Jacques	Neurochirurgie
M. CAMILLERI Lionel	Chirurgie Thoracique et Cardio-Vasculaire
M. DAPOIGNY Michel	Gastro-Entérologie

M.	LLORCA Pierre-Michel	Psychiatrie d'Adultes
M.	PEZET Denis	Chirurgie Digestive
M.	SOUWEINE Bertrand	Réanimation Médicale
M.	BOISGARD Stéphane	Chirurgie Orthopédique et Traumatologie
Mme	DUCLOS Martine	Physiologie
M.	SCHMIDT Jeannot	Médecine d'Urgence
M.	BERGER Marc	Hématologie
M.	GARCIER Jean-Marc	Anatomie-Radiologie et Imagerie Médicale
M.	ROSSET Eugénio	Chirurgie Vasculaire
M.	SOUBRIER Martin	Rhumatologie
M.	ABERGEL Armando	Hépatologie
Mlle	BARTHELEMY Isabelle	Chirurgie Maxillo-Faciale
M.	RUIVARD Marc	Médecine Interne

**PROFESSEURS DE  
1ère CLASSE**

M.	CAILLAUD Denis	Pneumo-physiologie
M.	VERRELLE Pierre	Radiothérapie option Clinique
M.	D'INCAN Michel	Dermatologie -Vénérologie
Mme	JALENQUES Isabelle	Psychiatrie d'Adultes
M.	GERBAUD Laurent	Epidémiologie, Economie de la Santé et Prévention
M.	TAUVERON Igor	Endocrinologie et Maladies Métaboliques
M.	MOM Thierry	Oto-Rhino-Laryngologie
M.	RICHARD Ruddy	Physiologie
M.	SAPIN-DEFOUR Vincent	Biochimie et Biologie Moléculaire
M.	BAY Jacques-Olivier	Cancérologie
M.	COUDEYRE Emmanuel	Médecine Physique et de Réadaptation
Mme	GODFRAIND Catherine	Anatomie et Cytologie Pathologiques
M.	LAURICHESSE Henri	Maladies Infectieuses et Tropicales
M.	TOURNILHAC Olivier	Hématologie
M.	CHIAMBARETTA Frédéric	Ophthalmologie
M.	FILAIRE Marc	Anatomie – Chirurgie Thoracique et Cardio-Vasculaire
M.	GALLOT Denis	Gynécologie-Obstétrique
M.	GUY Laurent	Urologie
M.	TRAORE Ousmane	Hygiène Hospitalière
M.	ANDRE Marc	Médecine Interne
M.	BONNET Richard	Bactériologie, Virologie
M.	CACHIN Florent	Biophysique et Médecine Nucléaire
M.	COSTES Frédéric	Physiologie
M.	FUTIER Emmanuel	Anesthésiologie-Réanimation
Mme	HENG Anne-Elisabeth	Néphrologie
M.	MOTREFF Pascal	Cardiologie
Mme	PICKERING Gisèle	Pharmacologie Clinique
M.	RABISCHONG Benoît	Gynécologie Obstétrique
M.	CHABROT Pascal	Radiologie et Imagerie Médicale
M.	DESCAMPS Stéphane	Chirurgie Orthopédique et Traumatologique
Mme	HENQUELL Cécile	Bactériologie Virologie
M.	POMEL Christophe	Cancérologie – Chirurgie Générale

**PROFESSEURS DE  
2ème CLASSE**

Mme CREVEAUX Isabelle	Biochimie et Biologie Moléculaire
M. FAICT Thierry	Médecine Légale et Droit de la Santé
Mme KANOLD LASTAWIECKA Justyna	Pédiatrie
M. TCHIRKOV Andréï	Cytologie et Histologie
M. CORNELIS François	Génétique
M. LESENS Olivier	Maladies Infectieuses et Tropicales
M. AUTHIER Nicolas	Pharmacologie Médicale
M. BROUSSE Georges	Psychiatrie Adultes/Addictologie
M. BUC Emmanuel	Chirurgie Digestive
M. LAUTRETTE Alexandre	Néphrologie Réanimation Médicale
Mme BRUGNON Florence	Biologie et Médecine du Développement et de la Reproduction
M. ESCHALIER Romain	Cardiologie
M. MERLIN Etienne	Pédiatrie
Mme TOURNADRE Anne	Rhumatologie
M. DURANDO Xavier	Cancérologie
M. DUTHEIL Frédéric	Médecine et Santé au Travail
Mme FANTINI Maria Livia	Neurologie
M. SAKKA Laurent	Anatomie – Neurochirurgie
M. BOURDEL Nicolas	Gynécologie-Obstétrique
M. GUIEZE Romain	Hématologie
M. POINCLOUX Laurent	Gastroentérologie
M. SOUTEYRAND Géraud	Cardiologie
M. EVRARD Bertrand	Immunologie
M. POIRIER Philippe	Parasitologie et Mycologie
Mme PHAM DANG Nathalie	Chirurgie Maxillo-Faciale et Stomatologie
Mme SARRET Catherine	Pédiatrie

**PROFESSEURS DES UNIVERSITES**

M. CLEMENT Gilles	Médecine Générale
Mme MALPUECH-BRUGERE Corinne	Nutrition Humaine
M. VORILHON Philippe	Médecine Générale

**PROFESSEURS ASSOCIES DES UNIVERSITES**

Mme BOTTET-MAULOUBIER Anne	Médecine Générale
M. CAMBON Benoît	Médecine Générale
M. TANGUY Gilles	Médecine Générale



**MAITRES DE CONFERENCES DES UNIVERSITES -  
PRATICIENS HOSPITALIERS**

**MAITRES DE CONFERENCES  
HORS CLASSE**

Mme CHAMBON Martine	Bactériologie Virologie
Mme BOUTELOUP Corinne	Nutrition

**MAITRES DE CONFERENCES DE  
1<sup>ère</sup> CLASSE**

M. MORVAN Daniel	Biophysique et Traitement de l'Image
Mlle GOUMY Carole	Cytologie et Histologie, Cytogénétique
Mme FOGLI Anne	Biochimie Biologie Moléculaire
Mlle GOUAS Laetitia	Cytologie et Histologie, Cytogénétique
M. MARCEAU Geoffroy	Biochimie Biologie Moléculaire
Mme MINET-QUINARD Régine	Biochimie Biologie Moléculaire
M. ROBIN Frédéric	Bactériologie
Mlle VERONESE Lauren	Cytologie et Histologie, Cytogénétique
M. DELMAS Julien	Bactériologie
Mlle MIRAND Audrey	Bactériologie Virologie
M. OUCHCHANE Lemlih	Biostatistiques, Informatique Médicale et Technologies de Communication
M. LIBERT Frédéric	Pharmacologie Médicale
Mlle COSTE Karen	Pédiatrie
Mlle AUMERAN Claire	Hygiène Hospitalière
Mme CASSAGNES Lucie	Radiologie et Imagerie Médicale
M. LEBRETON Aurélien	Hématologie
M. BUISSON Anthony	Gastroentérologie

**MAITRES DE CONFERENCES DE  
2<sup>ème</sup> CLASSE**

Mme PONS Hanaë	Biologie et Médecine du Développement et de la Reproduction
M. JABAUDON-GANDET Matthieu	Anesthésiologie – Réanimation Chirurgicale
M. BOUVIER Damien	Biochimie et Biologie Moléculaire
M. COLL Guillaume	Neurochirurgie
M. MAQDASY Salwan	Endocrinologie, Diabète et Maladies Métaboliques
Mme NOURRISSON Céline	Parasitologie - Mycologie
M. GODET Thomas	Anesthésiologie-Réanimation et Médecine Péri-Opératoire
M. LACHAL Jonathan	Pédopsychiatrie
M. MOUSTAFA Farès	Médecine d'Urgence

## MAITRES DE CONFERENCES DES UNIVERSITES

Mme VAURS-BARRIERE Catherine	Biochimie Biologie Moléculaire
M. BAILLY Jean-Luc	Bactériologie Virologie
Mle AUBEL Corinne	Oncologie Moléculaire
M. BLANCHON Loïc	Biochimie Biologie Moléculaire
Mle GUILLET Christelle	Nutrition Humaine
M. BIDEY Yannick	Oncogénétique
M. MARCHAND Fabien	Pharmacologie Médicale
M. DALMASSO Guillaume	Bactériologie
M. SOLER Cédric	Biochimie Biologie Moléculaire
M. GIRAUDET Fabrice	Biophysique et Traitement de l'Image
Mme VAILLANT-ROUSSEL Hélène	Médecine Générale
Mme LAPORTE Catherine	Médecine Générale
M. LOLIGNIER Stéphane	Neurosciences – Neuropharmacologie
Mme MARTEIL Gaëlle	Biologie de la Reproduction
M. PINEL Alexandre	Nutrition Humaine
M. PIZON Frank	Santé Publique

## MAITRES DE CONFERENCES ASSOCIES DES UNIVERSITES

M. BERNARD Pierre	Médecine Générale
Mme ESCHALIER Bénédicte	Médecine Générale
Mme RICHARD Amélie	Médecine Générale
M. TESSIERES Frédéric	Médecine Générale
Mme ROUGE Laure	Médecine Générale

# TABLE DES MATIERES

LISTE DES TABLEAUX ET FIGURES .....	9
CITATIONS.....	10
REMERCIEMENTS .....	11
PRÉAMBULE.....	12
1. INTRODUCTION .....	13
2. MATÉRIEL .....	15
3. MÉTHODES.....	16
4. RÉSULTATS .....	20
4.1. Échantillonnage.....	20
4.2. Analyse descriptive .....	21
4.3. Corrélation.....	24
5. DISCUSSION .....	25
5.1. Discussion de la méthodologie.....	25
5.2. Discussion sur les résultats.....	29
6. CONCLUSION.....	30
Annexe I. Cartographie de la région Auvergne-Rhône-Alpes des taux de mortalité standardisés sur l'âge entre 2006 et 2013. ....	31

Annexe II. Grille de saisie distribuée aux experts.....32

Annexe III. Fréquence des différences du score de risque global entre l’algorithme et les experts.  
.....33

Annexe IV. Fréquence de la somme des différences du score de risque global entre l’algorithme  
et les experts. ....34

BIBLIOGRAPHIE .....35

SERMENT D'HIPPOCRATE.....36

RÉSUMÉ.....

## LISTE DES TABLEAUX ET FIGURES

### TABLEAUX

Tableau I	: Comparaison entre la population du pilote CEPPPIA et l'échantillon d'Algo-CEPPPIA.	20
Tableau II	: Répartition des scores de risque global de l'ensemble des experts par rapport à l'algorithme.	21
Tableau III	: Distribution du score de risque global des avis majoritaires selon celui attribué par l'algorithme.	23

### FIGURES

Figure 1	: Attribution du risque global par l'algorithme selon les sous catégories.	15
Figure 2	: Méthode d'échantillonnage des fiches du pilote CEPPPIA pour Algo-CEPPPIA.	17
Figure 3	: Répartition des scores de risque global selon l'expert, l'algorithme et la majorité.	21
Figure 4	: Matrice de corrélation de Spearman deux à deux sur le score de risque global entre les experts, la majorité ainsi que l'algorithme.	23

## CITATIONS

*« Quand quelqu'un désire la santé, il faut d'abord lui demander s'il est prêt à supprimer les causes de sa maladie. Alors seulement est-il possible de l'aider. »*

Hippocrate

*« Avec une machine qui n'éprouve pas d'empathie, le phénomène de transfert psychologique ne peut se faire. Nous allons vers un monde extrêmement technique, avec des algorithmes extraordinaires mais nous aurons toujours besoin d'un médecin. »*

Dr Loïc ETIENNE, urgentiste à Paris

*« I know a man who gave up smoking, drinking, sex, and rich food. He was healthy right up to the day he killed himself. »*

Johnny CARSON, humoriste américain

## REMERCIEMENTS

**Monsieur le Professeur Laurent GERBAUD**  
**Professeur des Universités et Praticien Hospitalier**  
**Service de Santé Publique du CHU et Service de Santé Universitaire**

Je vous remercie de m'avoir fait l'honneur d'accepter de présider ce jury.

Merci Laurent de m'avoir aiguillé vers ce terrain de stage, qui en plus d'aboutir à ce travail, m'a permis d'aborder le monde de la recherche qui m'était alors inconnu. Je ne vous serai jamais assez reconnaissant de l'attention bienveillante, du soutien inconditionnel, de votre disponibilité et de la prévenance de tous les instants dont vous avez fait preuve à mon égard au cours de mon internat. Ce fut un réel plaisir et un privilège d'être accompagné par vous en tant que représentant de la spécialité au cours de ces quatre dernières années.

**Monsieur le Professeur Ruddy RICHARD**  
**Professeur des Universités et Praticien Hospitalier**  
**Centre de Recherche en Nutrition Humaine**

Merci de m'avoir permis de participer à ce grand projet qu'est le projet CEPPPPIA. Ce semestre riche et varié m'a donné l'opportunité de constater à quel point s'investir dans un projet de recherche pouvait être stimulant. Merci pour l'accompagnement et la réactivité dont vous avez fait preuve en tant que directeur de thèse.

**Monsieur le Professeur Jacques DAUPLAT**  
**Professeur des Universités et Praticien Hospitalier Honoraire**  
**Centre Jean PERRIN**

Un merci particulier au Pr. Jacques DAUPLAT, pour votre humanité et volontarisme. Votre persévérance et l'optimisme que vous diffusez sont un modèle dont tout le monde devrait s'inspirer. Ce fut un réel honneur d'avoir pu échanger aussi librement avec un professionnel de votre envergure, au parcours aussi exceptionnel.

**Monsieur le Docteur Bruno PEREIRA**  
**Docteur en Biostatistiques**  
**Direction de la Recherche Clinique et de l'Innovation**

Merci Dr. Bruno PEREIRA pour votre pédagogie et disponibilité. C'est un réel plaisir de travailler à vos côtés, et ce depuis notre première rencontre. Je ne peux vous cacher mon enthousiasme à réaliser mon premier semestre de docteur junior au sein de votre équipe.

### A l'équipe de l'étude Algo-CEPPPPIA

Un grand MERCI à vous Pr. DURANDO, Pr. RANNOU, Pr. SOUTEYRAND, Dr. CHABANAS, Dr. FARIGON et Dr. GAY-BELLILE qui avez composé le comité d'expert. Merci pour votre temps et engagement, c'est grâce à vous que cette étude a pu être réalisée.

Sans oublier Lynda, mon duo d'un semestre, ainsi que toute l'équipe du CRNH, que je remercie pour votre accueil chaleureux et votre professionnalisme.



Ruddy Richard

## PRÉAMBULE

Le projet CEPPPIA (Centre Expérimental de Prévention Prédicative Participative et Individualisée en Auvergne) a vu le jour en 2014 lors d'une commission « santé et bien-être » organisée par le conseil de développement du Grand Clermont, présidée par le Pr Jean Marc MONTEIL. Les porteurs initiaux du projet ont été le Pr Jean CHAZAL, le Pr Jacques DAUPLAT ainsi que la Pr Pascale DUCHÉ. CEPPPIA initialement porté par la DRCI du CHU de Clermont-Ferrand, a intégré la structure du CRNH d'Auvergne (Centre de Recherche en Nutrition Humaine) courant 2019, le Pr Ruddy RICHARD en a assuré la direction médicale et scientifique. Il a été lancé dans le but premier de bâtir un service public fiable s'inspirant de la médecine 4P.

Ce nouveau concept de médecine a été présenté en 2013 aux États-Unis par le biologiste Leroy HOOD. Il se base sur une **P**révention primaire avec une prise en charge **P**ersonnalisée, en impliquant la **P**articipation active du sujet. Elle se veut **P**rédictive en calculant le risque encouru de développer une maladie selon une cartographie précise et personnalisée.

Afin de prédire ce risque, le Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes (LIMOS), en collaboration avec le conseil scientifique de CEPPPIA, a développé un algorithme. Il a nécessité la création d'un comité scientifique constitué de 13 professionnels médicaux et méthodologistes issus du CHU de Clermont-Ferrand et du Centre Jean PERRIN<sup>1</sup>. Il se base sur une littérature scientifique actualisée, renforcée par un système d'auto-apprentissage (machine learning) développé par le LIMOS. Le questionnaire résultant est constitué de 231 items.

---

<sup>1</sup> J.Dauplat, J.Chazal, P.Duché, R.Richard, L.Gerbaut, Y.J.Bignon, Y.Boirie, G.Brousse, A.Chamoux, G.Clément, J.R.Lusson, F.Toumani, B.Pereira



Par la suite, une étude pilote a été réalisée de décembre 2017 à mai 2018, validée par le Comité de protection des personnes (CPP) et la Commission nationale de l'informatique et des libertés (CNIL). Un lien donnant accès au questionnaire en ligne a été mis à disposition. Sur les 3112 « clicks » de consultations, 570 inscriptions ont été réalisées, pour finalement aboutir à 250 Fiches de Santé Individuelle (FSI), aboutissant à un taux de réponse final de 8%. Les participants étaient volontaires et non rémunérés, âgés de 35 à 55 ans. Ces FSI constituent la source de données utilisées pour cette étude.

## **1. INTRODUCTION**

La France fait partie des pays où l'espérance de vie à la naissance est parmi les plus élevées au sein de l'Union Européenne. Cependant, elle présente un taux important de mortalité prématurée avant 65ans (1). Les principales causes de décès relèvent de maladies chroniques, les cancers et maladies cardio-vasculaires arrivant en tête (2). Leurs incidences et prévalences ne cessent d'augmenter, malgré une diminution de la mortalité standardisée sur l'âge depuis deux décennies.

Les facteurs comportementaux identifiés tels que la consommation de toxiques, l'alcool ou le tabac sont en baisse sur les dix dernières années, mais restent bien supérieurs à la moyenne européenne (1). La prévalence de l'obésité ainsi que la sédentarité sont en hausse sur l'ensemble des pays Européens (3).

L'incidence de ces pathologies chroniques et des comportements à risque est plus importante dans le nord et le centre de la France. L'ex-Auvergne présente un taux de mortalité prématuré supérieur à la moyenne nationale (Annexe I), sans que des facteurs de risques particuliers soient identifiés.

Reconnue pour ses compétences médicales, la France réoriente depuis 2016 sa politique de santé publique vers la prévention (4). Les moyens financiers attribués restent cependant maigres malgré une constante augmentation. Ils représentaient 2,4% de la consommation de soins et de biens médicaux (CSBM) en 2019 contre 4,9% en 2020 (5).

Le projet CEPPPIA répond à cette problématique de Santé Publique contemporaine. Son algorithme intègre l'ensemble des paramètres connus à ce jour afin de déterminer un niveau de risque global à contracter une pathologie chronique.

Il est nécessaire de valider cet algorithme en confrontant les résultats qu'il a généré au cours de l'étude pilote avec de nouveaux médecins « naïfs » afin de s'assurer de sa robustesse (6). La prédiction de ce score de risque global ayant valeur de diagnostic médical, il importe de s'assurer de la meilleure restitution possible, étant entendu qu'un diagnostic réalisé à tort ou une sous-évaluation des facteurs de risque peut nuire au sujet.

Algo-CEPPPIA vise à mettre en relation la prédiction du risque global de développer une pathologie chronique générée par l'algorithme à celle fournie par un comité d'experts médicaux, considéré ici comme « gold standard ». L'objectif principal est d'estimer et de mesurer la corrélation entre le risque prédit par l'algorithme et le groupe des experts médicaux. L'objectif secondaire est d'évaluer la cohérence entre les avis des experts.

## 2. MATÉRIEL

Les 250 FSI utilisées sont issues de l'étude pilote CEPPPIA, soumise au CPP<sup>2</sup> ainsi qu'à la CNIL<sup>3</sup>.

La population étudiée est âgée de 35 à 55 ans. Nous avons également utilisé le score de risque global de ces fiches, généré par l'algorithme. Ce dernier est issu d'une matrice combinant les risques pathologiques et comportementaux. Les pathologies étudiées sont du domaine cancérologique, cardiovasculaire et métabolique. Les comportements à risques sont dans les champs de la nutrition, de l'addiction, et de l'activité physique/sédentarité (Figure 1).

Le risque global prend en compte trois catégories :

- G0 quand il n'existe pas de risque identifiable
- G1 si un risque modéré relevant de la prévention primaire a été identifié
- G2 lorsque le risque était élevé, nécessitant une prise en charge médicale

Comportement \ Pathologie	P0	P1	P2
C0	G0	G1	G2
C1	G1	G1	G2
C2	G1	G2	G2

Pathologie (P)									Comportement (C)								
Cancer			Cardiovasculaire			Métabolique			Activité physique et sédentarité			Nutrition			Addictions		
R0	R1	R2	R0	R1	R2	R0	R1	R2	R0	R1	R2	R0	R1	R2	R0	R1	R2

**Figure 1** : Attribution du risque global par l'algorithme selon les sous catégories. Le code couleur permet de distinguer les niveaux de risque avec G0 en vert, G1 en orange et G2 en rouge. Le risque global est issu de la matrice entre le risque pathologique (P) et le risque comportemental (C). Les scores P et C sont chacun définis selon 3 sous-catégories de risque présentant 3 grades. R0 pour l'absence de risque, R1 pour un risque modéré et R2 pour un risque élevé.

<sup>2</sup> CPP SOOM II, référence : ID-RCB n°2017-A00666-47, Toulouse, France

<sup>3</sup> CNIL n°0189 au CHU de Clermont-Ferrand ; NCT03378895 sur ClinicalTrials.gov

Le comité d'experts pour la réalisation de cette étude est « naïf » du projet CEPPPIA. Ces derniers n'ont pris part ni à la constitution initiale de l'algorithme ni à la réflexion autour de cette étude. Ils ont été choisis afin de représenter les différentes spécialités en relation avec les pathologies et les facteurs de risques concernés. Le groupe se compose de six médecins ayant pour spécialité : la cardiologie, la cancérologie, l'oncogénétique, la médecine générale, les maladies métaboliques, la nutrition ainsi que la médecine du sport.

### **3. MÉTHODES**

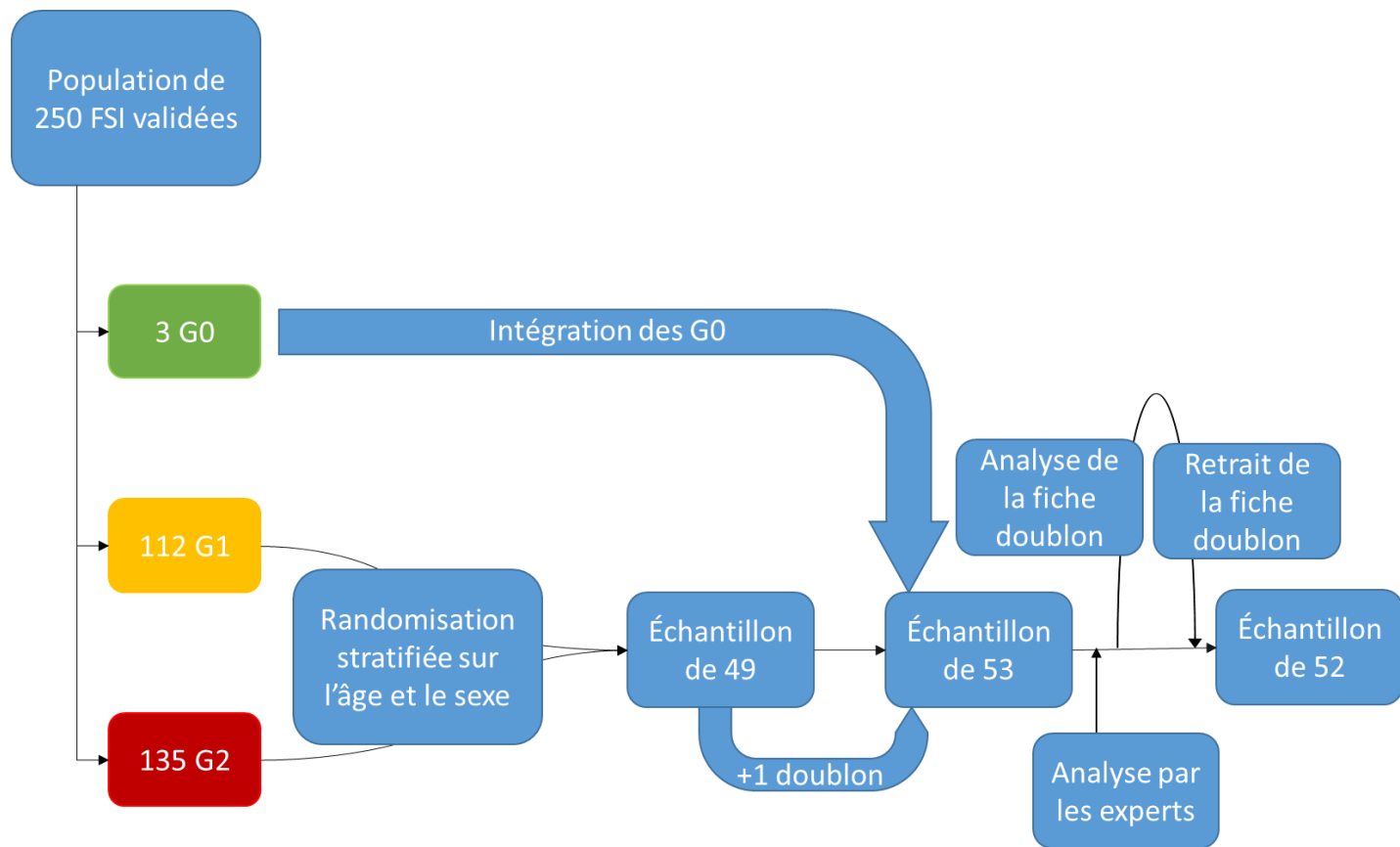
Il s'agit d'une étude de non- infériorité de l'algorithme CEPPPIA par rapport aux résultats obtenus avec l'avis majoritaire d'un nouveau comité de six experts « naïfs », constituant notre gold standard. Le critère de jugement principal porte sur la concordance des scores entre l'algorithme et le comité d'expert, avec un coefficient de corrélation du risque global, dont le seuil est fixé  $r \geq 0,7$ .

L'estimation du nombre de fiches nécessaires a été réalisée au travers de la littérature scientifique (7). Afin d'avoir une puissance suffisante ( $\beta = 20\%$ ) avec un risque d'erreur  $\alpha \leq 5\%$ , 50 observations étaient requises et cet effectif restait compatible avec le travail demandé aux experts médicaux.

Afin de se focaliser sur les fiches et pouvoir apprécier au mieux la variabilité entre les médecins, chaque expert a analysé l'ensemble des fiches.

Le temps d'analyse d'une fiche par un expert a été estimé à 15 minutes à la suite de la participation de deux médecins volontaires.

L'échantillonnage a été randomisé et stratifié selon les critères de l'âge, du sexe ainsi que du score de risque global. Sur les 250 fiches initiales, l'algorithme a classé seulement 3 d'entre elles comme présentant un risque global égal à 0. Il a été convenu d'exclure ces observations avant l'échantillonnage, puis de les rajouter en plus des fiches randomisées. L'échantillon obtenu était de 49 fiches, auxquelles il a été ajouté les 3 fiches cotées G0 par l'algorithme ainsi qu'une fiche en doublon (Figure 2).



**Figure 2 :** Méthode d'échantillonnage des fiches du pilote CEPPPPIA pour Algo-CEPPPPIA. Un échantillon de 52 fiches réalisé parmi la population de 250 issues de l'étude pilote, stratifiée selon le groupe de risque global, l'âge et le sexe. Les fiches cotées G0 par l'algorithme ont été rajoutées à l'échantillon après échantillonnage. Une fiche incluse en doublon a été supprimée avant le traitement des données.

La présentation du projet ainsi que les modalités de renseignement des résultats ont été réalisées au cours de trois réunions.

Le protocole suivant a été mis en œuvre.

Pour chaque fiche, l'expert devait en premier lieu juger le niveau de risque global, puis dans un second temps, détailler les niveaux de risques selon les catégories de pathologies (cancéreuse, cardiologique et métabolique) et de comportements à risque (sédentarité, nutrition et addiction). L'importance de l'ordre de saisie a été soulignée afin de recueillir le ressenti global du médecin sur le cas traité, sans qu'il ait été guidé par la cotation plus fine des sous-scores. Le jugement holistique du médecin est ainsi récupéré via le score de risque global.

La durée d'analyse pour les 53 fiches (doublet inclus) a été estimée à 13 heures, ce qui a écarté la possibilité d'une analyse en présentiel. Ces dernières ont été anonymisées et envoyées en deux lots, l'un de 27 fiches et l'autre de 26 fiches, via un hébergement sécurisé. Un fichier Excel structuré leur a été transmis afin d'uniformiser la saisie des données (Annexe II). La réception des avis individuels s'est faite au fil de l'eau via la messagerie sécurisée du CHU de Clermont-Ferrand sur une adresse spécifique au projet.

Une fois l'ensemble des fichiers réceptionné, l'avis des experts a été exprimé en un avis majoritaire selon la logique du risque maximal. Ainsi, en cas d'égalité du nombre d'attribution du score de risque global entre deux modalités, c'est le risque le plus péjoratif qui a été considéré.

Une analyse descriptive des données a été réalisée dans un premier temps, permettant d'apprécier la cohérence des scores entre chaque expert, ainsi que la dispersion globale de la différence entre l'algorithme et les experts.

Dans un second temps, des statistiques inférentielles ont été réalisées sur le score de risque global par des corrélations de Spearman deux à deux entre les avis des experts, l'avis majoritaire et l'algorithme. Le seuil de significativité considéré a été fixé à  $p \leq 0,05$ .

Les données ont été traitées sous R version 4.0.5 (2021-03-31) -- "Shake and Throw".

Il s'est écoulé un total de 4 mois entre la diffusion des premières fiches et la réception intégrale des 53 fiches (49 échantillonnées + 1 doublon + 3 G0). Une réunion de restitution des résultats a été organisée à la fin de l'étude.

## 4. RÉSULTATS

### 4.1.Échantillonnage

L'échantillon obtenu était représentatif de la population, homogène sur le score global de risque, le sexe, ainsi que la moyenne d'âge et son écart-type (Tableau I). Il n'y a pas eu de variable non attribuée. L'observation en doublon a été retirée pour la réalisation des analyses, amenant notre échantillon final à 52 observations.

**Tableau I** : Comparaison entre la population du pilote CEPPPIA et l'échantillon d'Algo-CEPPPIA.

		Population (N=250)		Echantillon (n=52)	
		N (%)	$\sigma$	n (%)	$\sigma$
Sexe	Homme	90 (35%)		17 (33%)	
	Femme	160 (65%)		35 (67%)	
Age moyen		46,6	6,1	47,7	6,5
Risque global Gx	G0	3 (1%)		3 (6%)	
	G1	112 (45%)		22 (42%)	
	G2	135 (54%)		27 (52%)	

Les 3 G0 rajoutés post-échantillonnage représentaient 3 femmes, âgées en moyenne de 48,3 ans.

Au final, les 52 fiches uniques analysées par les 6 experts ainsi que l'algorithme ont abouti à une base de données complète de 364 observations.



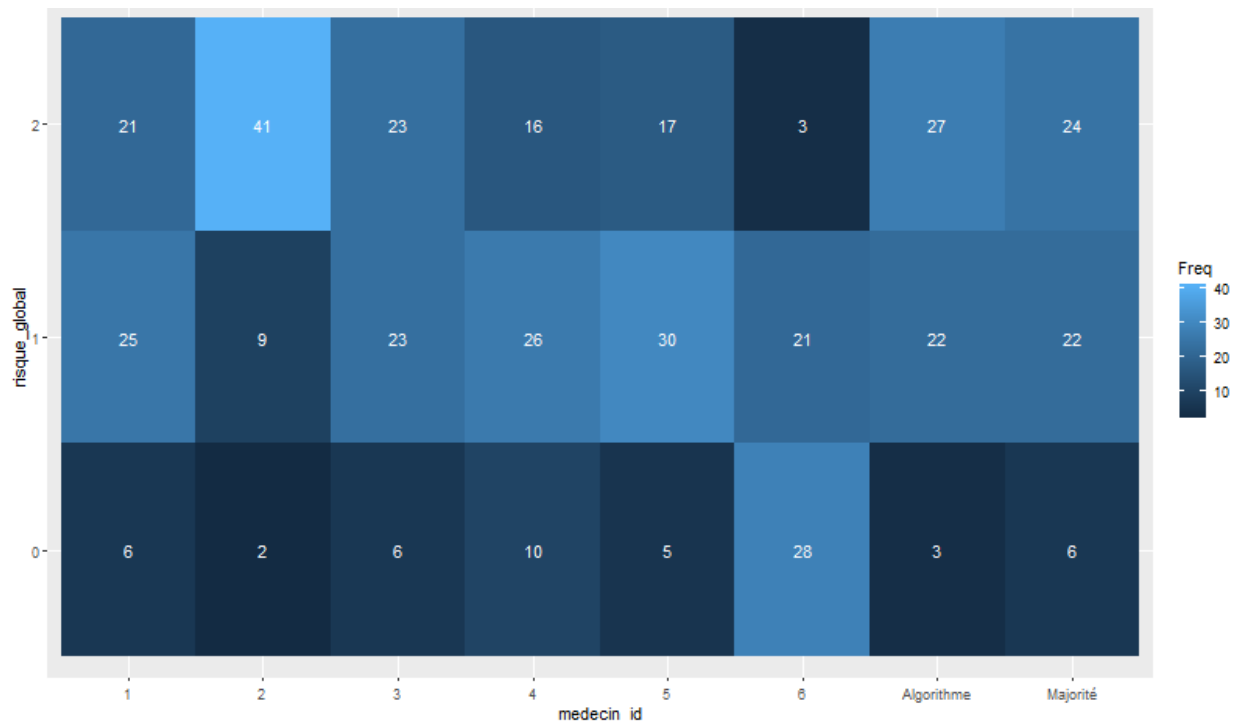
## 4.2. Analyse descriptive

La cotation des experts a été répartie de la manière suivante (Tableau II).

**Tableau II** : Répartition des scores de risque global de l'ensemble des experts par rapport à l'algorithme.

	Score de risque global n (%)		
	G0	G1	G2
Experts	57 (18%)	134 (43%)	121 (39%)
Algorithme	3 (6%)	22 (42%)	27 (52%)

Les experts ont présenté des cotations hétérogènes entre eux (Figure 3). Le médecin n°2 a attribué plus fréquemment des scores élevés G2 (n=41 ; 79%) à l'opposé du médecin n°6 qui a fréquemment attribué des G0 (n=28 ; 54%). Les 4 autres experts ont présenté une répartition similaire des scores de risque global, se rapprochant également de l'avis majoritaire et de l'algorithme.



**Figure 3** : Répartition des scores de risque global selon l’expert, l’algorithme et la majorité. Représente pour chaque expert (abscisse) la fréquence d’attribution des différents scores de risque global (ordonnée). L’intensité de réponse pour chaque colonne se traduit par un dégradé de couleur allant du noir (moins fréquent) au bleu clair (plus fréquent). La somme de chaque colonne correspond aux 52 fiches analysées par chaque expert.

Il a été apprécié la différence de cotation entre les experts et l’algorithme (Annexe III). L’algorithme est en accord pour 158 (50%) des observations des experts. Dans les cas de désaccords, l’algorithme a coté un risque plus élevé sur 113 observations (73%). À 12 reprises (4%), l’algorithme a présenté un score totalement opposé à celui d’un médecin.

La somme des différences entre l’algorithme et l’avis des experts a été réalisée (Annexe IV). Les valeurs extrêmes obtenues ont été de -7 pour une fiche et de +6 pour une autre, 96% (n=50) des fiches se situant entre -4 et +5.

L'avis des experts n'a été unanime que pour une seule fiche, qui a été cotée G2 en accord avec l'algorithme. La majorité absolue définie par score attribué par 4 experts ou plus, a été atteinte dans 50% des cas (n=26). Une égalité des avis a concerné 10 fiches (19%), la majorité ayant été attribuée selon la logique de risque maximal pour ces dernières.

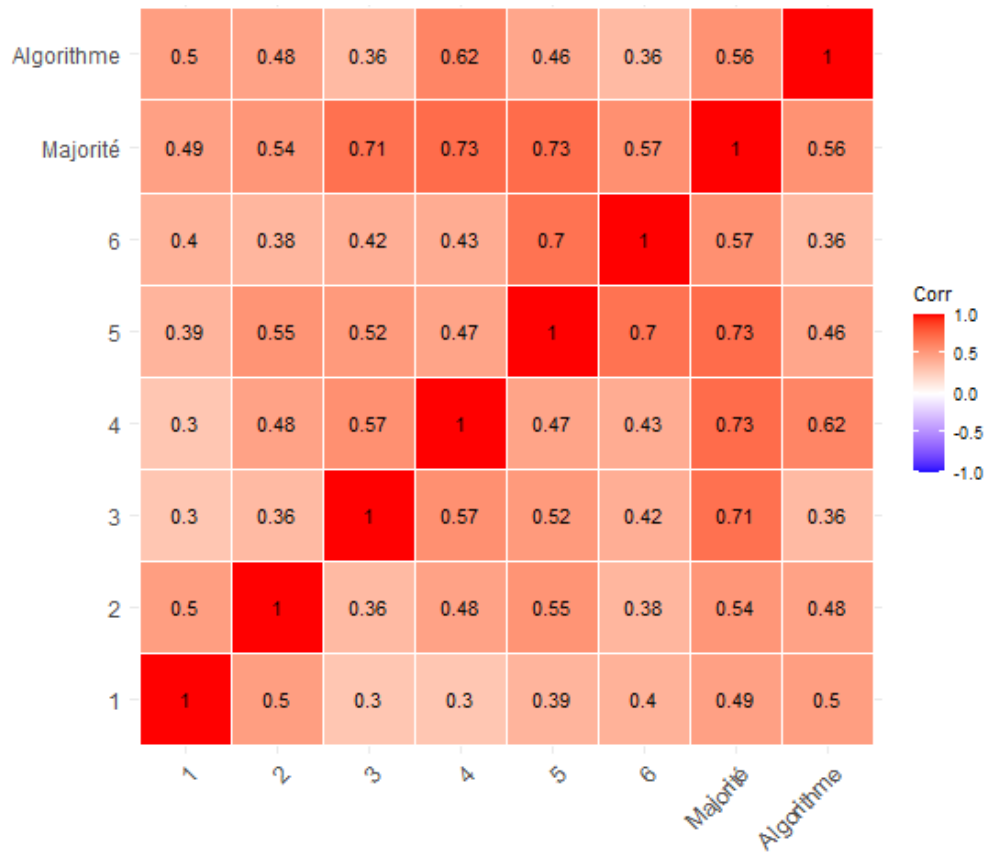
L'algorithme a été en accord avec la majorité des experts dans 62% des cas (n=32) (Tableau III). Aucune fiche n'a présenté de différence de 2 points entre l'avis majoritaire des experts et l'algorithme, dans 13 cas (25%) l'algorithme présentait un score supérieur de 1, et dans 7 cas (13%) un score inférieur de 1. Sur les 3 fiches G0 estimées par l'algorithme, une seule s'est révélée être en accord avec les experts.

**Tableau III** : Distribution du score de risque global des avis majoritaires selon celui attribué par l'algorithme.

Logique du risque maximum		Gx algorithme			
		G0	G1	G2	Total
Différence de Gx entre l'algorithme et l'avis majoritaires	-2	0	-	-	0
	-1	2	5	-	7
	0	1	12	19	32
	1	-	5	8	13
	2	-	-	0	0
	Total	3	22	27	52

### 4.3. Corrélation

Des corrélations de Spearman ont été réalisées deux à deux, comparant les avis des experts, l'avis majoritaire, ainsi que l'algorithme. L'ensemble des coefficients de corrélation ont été significativement positifs ( $p < 0,0001$ ), variant de 0,3 à 0,73 (Figure 4). Le coefficient de corrélation de l'avis majoritaire avec l'algorithme a été de 0,56 ( $p < 0,0001$ ), inférieur au seuil de 0,7 initialement fixé. L'écart-type résiduel s'est élevé à 0,63. Le coefficient de corrélation moyen entre l'algorithme et les experts a été de 0,46. Seule la corrélation entre 3 experts et l'avis majoritaire a atteint le seuil fixé de 0,7.



**Figure 4 :** Matrice de corrélation de Spearman deux à deux sur le score de risque global entre les experts, la majorité ainsi que l'algorithme. Présentation en miroir comparant les experts entre eux, en attribuant un coefficient de corrélation de Spearman pour chaque comparaison ( $p < 0,0001$ ). Un dégradé de couleur marque l'intensité de la corrélation, du rouge pâle pour la plus faible, au rouge foncé pour la plus forte. L'absence de bleu relève l'absence de corrélation négative.

## **5. DISCUSSION**

### **5.1. Discussion de la méthodologie**

Soumettre les mêmes fiches à chaque expert a eu le mérite de nous permettre d'analyser la variabilité entre ces derniers. Cela a néanmoins rendu impossible dans le même temps d'explorer l'intégralité des 250 fiches. Dès lors, nous avons décidé de procéder à un échantillonnage. Cette méthode a permis de mettre en avant que des cliniciens de spécialités différentes tiraient des conclusions hétérogènes sur l'état de santé d'une même personne. L'intégration d'une fiche en doublon a permis de confirmer l'assiduité des experts. Ces points ont relevé la difficulté à concevoir un algorithme faisant consensus.

Le choix d'intégrer l'intégralité des 3 fiches cotées G0 est également discutable. Il a été pris à la suite de la constatation étonnante de leur très faible proportion au sein de la population de fiches. Si nous avions voulu respecter la proportion de G0 au sein de l'échantillon, nous aurions eu deux cas de figures, soit une seule fiche, soit aucune. L'analyse de cette catégorie n'aurait alors eu aucune portée statistique, alors que l'intégration des 3 fiches concernées maximisait nos chances d'interprétation.

Initialement, l'étude présentait 3 phases d'analyse. Chaque phase constituait une étape comportant l'envoi des fiches, leur analyse et restitution par les experts, puis un traitement de la donnée.

La première phase constituée de 27 fiches, ainsi la seconde phase avec 26 fiches, étaient obligatoires dans cette étude, intégrant à elles deux le nombre de fiches minimales. La troisième phase était facultative. Si le coefficient de corrélation était proche de l'objectif fixé sans pour autant l'atteindre, elle permettait l'ajout de 9 fiches supplémentaires. Ce qui n'a pas été le cas avec un  $r = 0,56$  et  $p < 0,0001$ .

L'organisation de réunions a été envisagée à la fin de chaque phase, afin d'obtenir un avis de consensus selon la méthode Delphi. Cette dernière consiste à présenter anonymement les avis, permettant à chaque expert d'ajuster leur réponse au cours de plusieurs tours, afin d'obtenir une réponse consensuelle. Cependant, cette méthode n'était réalisable que si aucun biais d'estimation n'était repéré chez les experts.

Il a été jugé préférable de séparer en deux l'analyse de l'échantillon de 53 fiches et ceci pour deux raisons. La première était la difficulté à réunir de façon simultanée tous les experts. Il semblait plus simple de convoquer deux courtes réunions qu'une unique réunion plus longue. La seconde était le raccourcissement du délai entre la diffusion des fiches et la restitution des analyses. En effet, un envoi simultané de l'intégralité des 53 fiches aurait impliqué un temps d'analyse plus long, ce qui aurait pu induire un biais de mémoire, les cas étudiés étant plus nombreux et plus anciens. La méthodologie retenue participe de cette volonté de veiller à l'efficacité du traitement des dossiers, ainsi qu'à la qualité du potentiel consensus, ceci dans un double souci de réduire le temps de concertation et d'augmenter la faisabilité.

A l'issue de la première phase, il est apparu que l'expert n°6 présentait une tendance systématique à la sous-cotation par rapport à ses confrères. Il avait estimé 2 fiches à G2 (7%), alors que la moyenne des autres experts était à 14 G2 (52%). De la même façon, il présentait une notation moyenne de 0,74, alors que ses confrères étaient à 1,4. De ce fait, l'avis de consensus selon la méthode Delphi a été irréalisable et les réunions de concertation ont été annulées. Cette situation a mené au traitement des données selon la logique du risque maximal.

Il a été décidé de réunir des médecins experts dans leurs domaines afin d'obtenir la connaissance médicale la plus juste possible. Cette approche faisait sens si une discussion permettait de joindre les expertises de chacun afin d'arriver à un avis unique. Or, au cours de cette étude, seules les décisions individuelles ont été prises en compte. Au regard de ces éléments, il nous apparaît qu'il aurait été plus judicieux de solliciter des médecins généralistes, dont la profession permet une approche plus transversale du patient.

Le protocole d'étude a été réalisé afin de pallier aux freins principaux de réalisation que sont le temps demandé aux médecins pour analyser les fiches d'une part, et les modalités opérationnelles de traitement de l'information d'autre part.

Ainsi, les fiches ont été traitées à distance, les envois fragmentés, le temps total mis à disposition élargi à 4 mois, et la taille de l'échantillon réduite au maximum possible. Malgré ces mesures, il y a eu deux désistements et un nouveau recrutement d'experts avant le lancement de l'étude. Initialement prévus au nombre de 7, ils n'ont été que 6 à participer, générant de fait une problématique pour la définition de la majorité.

Cette étude a traité le score de risque global comme une variable quantitative continue. Or, celle-ci ne comporte que 3 modalités, issue d'une représentation clinique de l'individu au travers de la fiche. Le fait d'avoir un nombre aussi restreint de modalités a permis d'avoir une ségrégation claire des catégories entre elles.

Lors de la restitution de l'analyse des experts, une interprétation en sous catégories de risques pathologiques et comportementaux a été demandée. Non abordée dans ce travail, ils pourraient permettre d'élucider les différences d'avis selon la spécialité de l'expert. De même, chaque expert possédait une section afin de renseigner les commentaires sur chaque fiche. La présence de celle-ci était prévue afin de faciliter la mémorisation des cas lors des potentielles réunions. Aucune consigne de remplissage n'a été fournie, ne permettant pas de se baser sur ces informations. Leur analyse pourrait cependant éclairer certaines fiches présentant une forte discordance entre experts.

Il n'a été abordé que les pathologies cancéreuses, cardiaques et métaboliques. Les pathologies neurologiques, moins fréquentes que ces dernières, auraient pu figurer au sein de l'analyse (8).



## 5.2. Discussion sur les résultats

L'analyse descriptive des données a permis de suspecter une forte sous-cotation de l'expert n°6. Ce dernier présente pourtant des coefficients de corrélation se rapprochant de ceux de ses confrères. Le cas de l'expert n°1 nous met en garde sur la valeur à attribuer aux analyses descriptives. En effet, la répartition de la cotation des Gx semblait bien s'accorder avec la majorité de ses confrères. Cependant, il présente les moins bons coefficients de corrélation.

La somme des différences entre les médecins et l'algorithme présente des valeurs théoriques comprises entre -12 et +12. Ces dernières n'ont pas été atteintes, et 96% des fiches étaient situées entre -4 et +5, démontrant que les experts n'ont présenté aucun franc désaccord de cotation sur l'intégralité des fiches.

L'obtention d'un avis majoritaire a permis d'uniformiser les avis, et d'obtenir le risque global le plus probable. Il présente les meilleurs coefficients de corrélation aux praticiens, cependant, ceux-ci n'excèdent pas 0,73. Un avis de consensus aurait probablement présenté de meilleurs résultats sur ce point.

## 6. CONCLUSION

La corrélation entre l'avis majoritaire obtenu par la logique de risque maximum et l'algorithme a été de 0,56. Le seuil de corrélation initialement fixé à 0,7 n'ayant pas été atteint, nous ne pouvons pas affirmer que la prédiction réalisée par l'algorithme soit suffisamment corrélée à l'avis majoritaire des experts. Cette valeur seuil n'a été atteinte que lors de certaines corrélations d'experts avec l'avis majoritaire. Ce constat révèle que le seuil visé était élevé. La corrélation est classiquement estimée comme forte lorsque  $r > 0,5$ . Cependant, au regard de l'impact potentiel sur le sujet d'un diagnostic erroné, une bonne fiabilité de l'algorithme apparaît comme un impératif. La comparaison de l'algorithme avec l'avis majoritaire des experts selon une logique de risque maximal était en accord dans 62% des cas, et n'a présenté aucun désaccord extrême. Les cas divergents menaient deux fois plus fréquemment à une sur-cotation du risque par l'algorithme.

De fortes disparités entre les cotations des experts ont été constatées, pointant la complexité à prédire le risque de développer une pathologie chronique par un algorithme unique.

Il serait intéressant d'approfondir les résultats de cette étude en abordant les sous-catégories de risques, afin d'explorer les causes potentielles de ces divergences d'opinions. Il serait également envisageable de réaliser une nouvelle étude en recueillant un avis de consensus.

Le Doyen de l'UFR de Médecine,

Pr. Pierre CLAVELOU



Le Président du Jury,

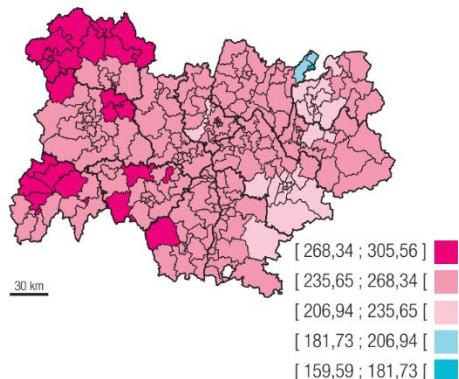
Pr. Laurent GERBAUD

C.H.U. de Clermont-Ferrand  
Service de Santé Publique  
Professeur Laurent GERBAUD  
7 place Henri Dunant  
63000 CLERMONT-FERRAND  
Tél : 04 73 750 340

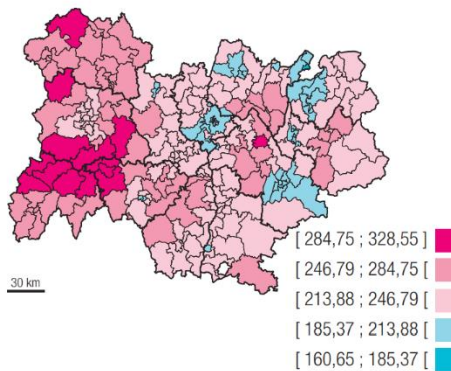
# Annexe I. Cartographie de la région Auvergne-Rhône-Alpes des taux de mortalité standardisés sur l'âge entre 2006 et 2013.

## Taux standardisé\* de mortalité 2006-2013 par cantons

### Cancers



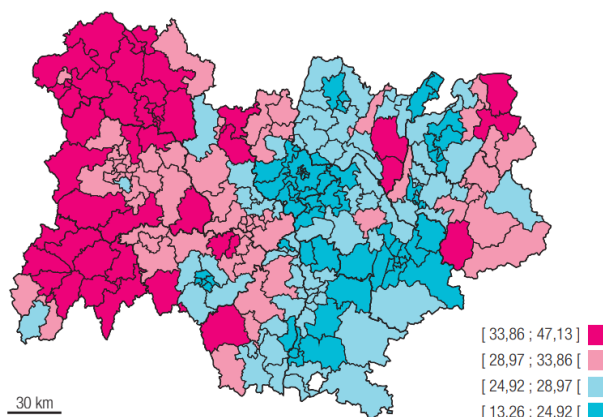
### Maladies cardio-vasculaires



## Taux standardisé\* de mortalité 2006-2013 par cantons

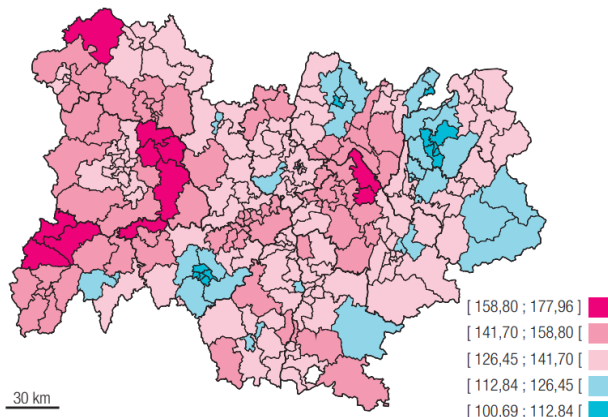
### Alcool

(Cancers des voies aérodigestives supérieures (VADS), cirrhoses du foie, psychoses alcooliques et alcoolisme)



### Tabac

(Cancers du poumon, cardiopathies ischémiques, bronchopneumopathies chroniques obstructives (BPCO))



\* Taux standardisé sur l'âge pour 100 000 habitants, selon la population de référence européenne (Revision of the European Standard Population, Eurostat 2013).

Sources • Certificats de décès (CépiDc) ; RP 2006 à 2012 (INSEE) ; exploitation FNORS.

## Annexe II. Grille de saisie distribuée aux experts.

	A	B	C	D	E	F	G	H	I	J
1	<b>Algo-CEPPPIA : Grille de saisie</b>									
3	Nom de l'expert :									
5	Phase	n° Fiche	Groupe de risque Global	Cancer	Cardio	Métabo	AP/S	Nutrition	Addiction	Commentaire
6	1	1	1	1	0	0	0	0	0	risque cutané
7	1	2	2	1	2	1	2	0	2	ORCV (Sexe, Cholest, Tabac, ATCD familial AVC, sédentari
8	1	3	1	0	0	0	0	1	0	risque cutané, ATCD cancers 1er degré, t bies sommeil,
9	1	4	2	2	1	1	0	0	2	
10	1	5	2	1	1	1	0	0	1	15 P/A, ATCD cancer chez père, alcoolisation aigue, tbles so
11	1	6	2	2	2	2	2	2	2	hiliaux, obésité (+HTA, hypercho, tabac actif ( <b>20 mois ou 2</b>
12	1	7	2	0	1	1	0	2	0	mation charcuterie-plats préparés-sucre-sels-rajout matières
13	1	8	2	2	1	1	0	2	0	cutané, ATCD tabac +++ sevré 50P/A, peu fruits-légumes.
14	1	9	2	2	1	1	1	1	2	Tabac, alcool, pdts sucrés
15	1	10	2	2	1	1	1	2	1	risque cutané (Uvs, oncie), thyroïde, cannabis, charcuterie-
16	1	11	2	2	1	2	1	1	0	risque cutané, colon, peu fruits
17	1	12	2	2	1	0	1	1	0	ue cutané, néoplasies 1er degré, pas sucres lents, sommeil-str
18	1	13	2	1	0	0	0	0	2	tabac
19	1	14	1	0	1	0	1	0	0	sommeil, tps de travail???
20	1	15	2	2	1	1	1	0	1	ATCD néoplasies
21	1	16	2	2	2	2	1	2	2	2 enfants bio/1 grossesse???
22	1	17	2	2	1	1	1	0	0	néoplasies 1er degré
23	1	18	0	0	0	0	0	0	0	
24	1	19	1	1	0	1	1	0	1	sédentarité, écrans
25	1	20	2	2	2	1	0	1	2	écrans, stress

Grille de saisie des résultats sous format Excel, utilisée pour la cotation des risques de l'échantillon d'Algo-CEPPPIA. Chaque expert a reçu un exemplaire de cette grille. La phase et le numéro de fiche ont été préremplis. Chaque risque a pu se voir attribuer une valeur numérique de 0 à 2. La zone de commentaire a été donnée à titre facultatif pour assister l'expert.

**Annexe III.** Fréquence des différences du score de risque global entre l’algorithme et les experts.

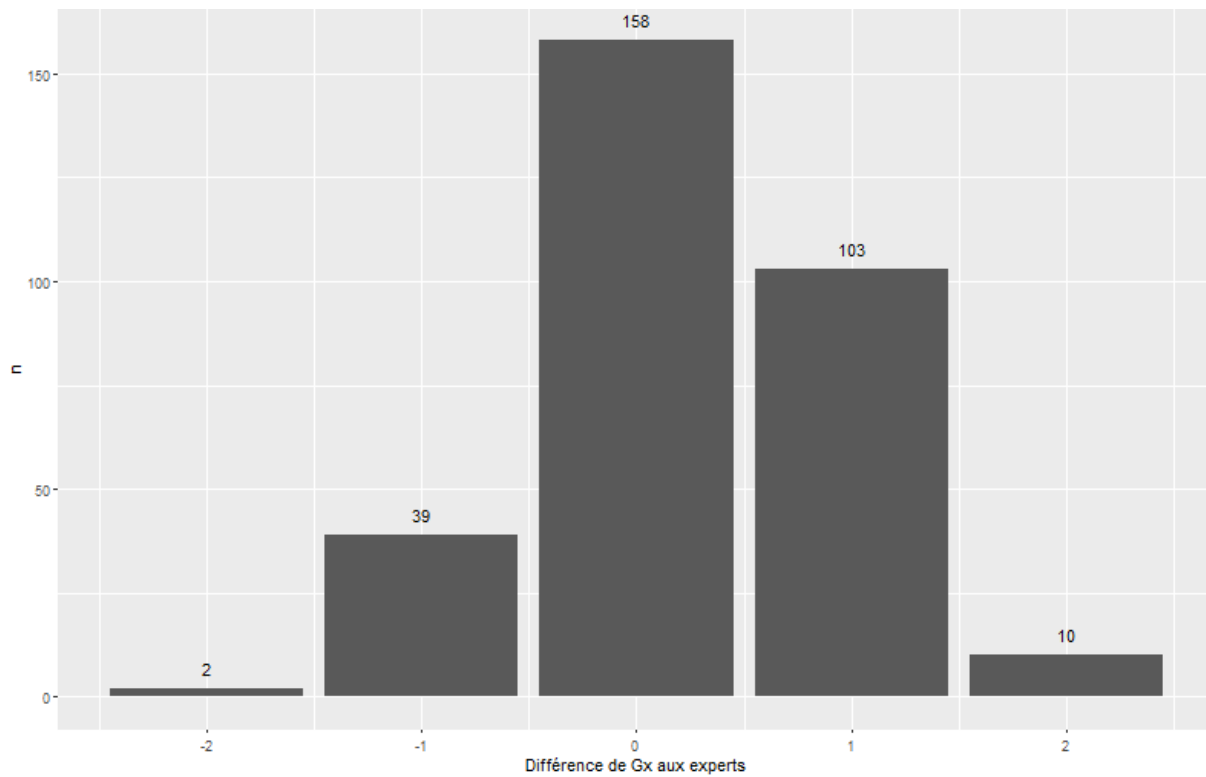


Diagramme en barre représentant la fréquence n (ordonnée) des différences des scores de risque global entre l’algorithme et chaque expert (abscisse). Présente un total de 312 observations (6 experts x 52 fiches) sur un intervalle de [-2 ; +2].

**Annexe IV.** Fréquence de la somme des différences du score de risque global entre l'algorithme et les experts.

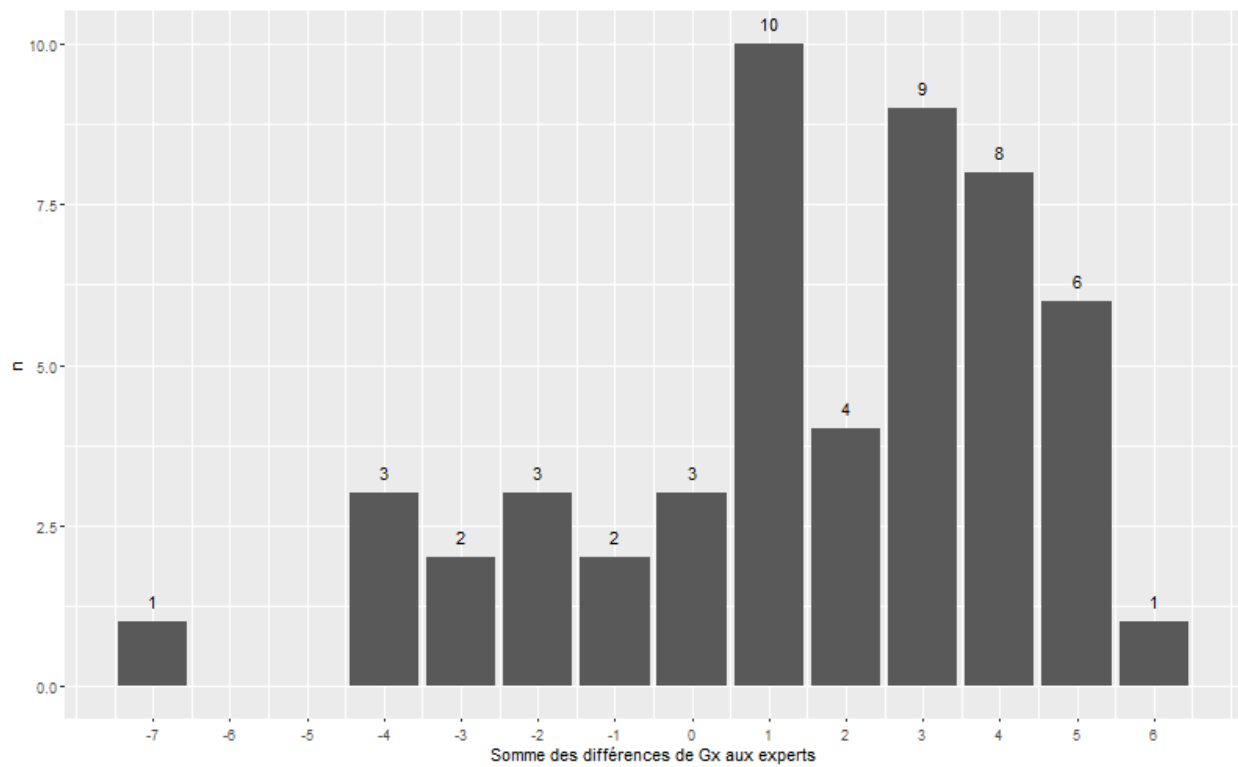


Diagramme en barre représentant la fréquence n (ordonnée) de la somme des différences de score de risque global entre l'algorithme et les experts (abscisse). Présente un total de 52 observations avec intervalle théorique de [-12 ; +12] ([-2 x 6 experts ; +2 x 6 experts]).

## BIBLIOGRAPHIE

1. Rapport-ESPF-2017.pdf [Internet]. [cité 5 oct 2021]. Disponible sur: <https://drees.solidarites-sante.gouv.fr/sites/default/files/2021-01/Rapport-ESPF-2017.pdf>
2. Espérance de vie en bonne santé – Indicateurs de richesse nationale | Insee [Internet]. [cité 3 mars 2020]. Disponible sur: <https://www.insee.fr/fr/statistiques/3281641?sommaire=3281778#consulter>
3. OECD Statistics [Internet]. [cité 29 mai 2020]. Disponible sur: <https://stats.oecd.org/>
4. Légifrance - Publications officielles - Journal officiel - JORF n° 0022 du 27/01/2016 [Internet]. [cité 6 oct 2021]. Disponible sur: [https://www.legifrance.gouv.fr/download/pdf?id=f1zqqKkO-FAUZH67\\_XjED1sDFihSq-tW46Kwa2ISZzs=](https://www.legifrance.gouv.fr/download/pdf?id=f1zqqKkO-FAUZH67_XjED1sDFihSq-tW46Kwa2ISZzs=)
5. Jabri K, Monirijavid S, Renaud T, Lenglard F, Polton D, Bauer-Eubriet V, et al. PANORAMAS DE LA SANTÉ - Les dépenses de santé en 2020 Résultats des comptes de la santé - DREES ÉDITION 2021. 2020;225.
6. Nagpal K, Foote D, Liu Y, Chen P-HC, Wulczyn E, Tan F, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *npj Digit Med*. 7 juin 2019;2(1):1-10.
7. Bujang MA, Baharum N. Sample Size Guideline for Correlation Analysis. *World Journal of Social Science Research*. 10 mars 2016;3:37.
8. Aouba A. Les causes médicales de décès en France en 2004 et leur évolution. :7.
9. PANORAMAS DE LA DREES SANTÉ - Les dépenses de santé en 2020 Résultats des comptes de la santé - ÉDITION 2021 [Internet]. [cité 6 oct 2021]. Disponible sur: <https://drees.solidarites-sante.gouv.fr/sites/default/files/2021-09/Les%20d%C3%A9penses%20de%20sant%C3%A9%20en%202020.pdf>

**(Conseil national de l'ordre des médecins)**

**SERMENT D'HIPPOCRATE**

Au moment d'être admis(e) à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité.

Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences.

Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences.

Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admis(e) dans l'intimité des personnes, je tairai les secrets qui me seront confiés. Reçu(e) à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs.

Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité.

Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré(e) et méprisé(e) si j'y manque.

MORTIER Romain Thomas,

Le 27/10/2021 à Clermont-Ferrand.



## SERMENT D'HIPPOCRATE

En présence des Maîtres de cette FACULTE et de mes chers CONDISCIPLES, je promets et je jure d'être fidèle aux lois de l'Honneur et de la Probité dans l'exercice de la Médecine.

Je donnerai mes soins gratuits à l'indigent et je n'exigerai jamais un salaire au-dessus de mon travail. Admis dans l'intérieur des maisons, mes yeux ne verront pas ce qui s'y passe, ma langue taira les secrets qui me seront confiés et mon état ne servira pas à corrompre les mœurs ni à favoriser le crime.

Respectueux et reconnaissant envers mes MAÎTRES, je rendrai à leurs enfants l'instruction que j'ai reçue de leurs pères.

Que les HOMMES m'accordent leur estime si je suis fidèle à mes promesses. Que je sois couvert d'OPPROBRE et méprisé de mes confrères si j'y manque.

MORTIER Romain Thomas,

Le 27/10/2021 à Clermont-Ferrand.

## **Algo CEPPPIA – Prédiction du risque de développement de maladies chroniques : validation d'un algorithme par corrélation avec un comité d'experts**

### **RÉSUMÉ**

#### **CONTEXTE**

La France est reconnue pour ses soins curatifs mais présente un volet préventif à développer. Le projet CEPPPIA se base sur la médecine 4P, proposant un programme de prévention personnalisé accessible au plus grand nombre. Pour ce faire, un questionnaire de santé informatisé très complet est analysé par un algorithme prédisant une gradation du risque de développer des pathologies chroniques.

#### **OBJECTIF**

Valider le niveau de risque global généré par l'algorithme en le confrontant à celui émanant de médecins spécialisés dans les domaines concernés : cardiologie, cancérologie, oncogénétique, endocrinologie & nutrition, activité physique, médecine générale.

#### **MÉTHODE**

Un échantillon de 52 questionnaires sur les 250 automatiquement scorés de 0 à 2 par l'algorithme lors de l'étude pilote a été analysé par chacun des 6 médecins « naïfs ». La mise en relation a été réalisée via un test de concordance, sur données répétées, entre le risque global donné par les médecins et l'algorithme, avec pour objectif un coefficient de corrélation  $\geq 0,7$ .

#### **RÉSULTATS**

Le coefficient de corrélation obtenu est de  $0,56 < 0,7$ , avec un écart-type résiduel de 0,63.

#### **CONCLUSION**

N'atteignant pas le seuil de corrélation initialement fixé, nous ne pouvons pas conclure que l'algorithme est fidèle à l'expertise médicale selon les conditions de cette étude.

#### **Mots-clés :**

- Médecine 4P
- Prévention primaire
- CEPPPIA
- Maladie chronique
- Algorithme prédictif
- Machine learning

07.10.2014  
C.H.U. de Clermont-Ferrand  
Service de Santé Publique  
Professeur Laurent GERBAUD  
7 place Henri Dunant  
63000 CLERMONT-FERRAND  
Tél : 04 73 750 340