

UNIVERSITÉ DE STRASBOURG

FACULTÉ DE MÉDECINE, MAÏEUTIQUE ET SCIENCES DE LA SANTÉ

ANNEÉ : 2023

N° : 006

THÈSE

PRÉSENTÉE POUR LE DIPLÔME D'ÉTAT

DE DOCTEUR EN MÉDECINE

Diplôme d'État

Mention : Médecine Générale

PAR

ZOLGER Michaël

Né le 13 février 1990, à Sarrebourg (57)

Analyse descriptive des sujets de thèses de médecine générale en France en 2021 et leurs évolutions
depuis 2015 en utilisant des outils de fouilles de texte.

Président de thèse

Monsieur le Professeur PELACCIA Thierry

Directeur de thèse

Monsieur le Docteur FABACHER Thibaut, Assistant Hospitalier Universitaire au service de Santé

Publique des Hôpitaux Universitaires de Strasbourg

B3 - MAITRES DE CONFERENCES DES UNIVERSITES (monoappartenant)

M. K. SOULIER	Département d'Histologie, Anatomie Médicale, Faculté de Médecine	72	La cardiologie - Histoire des sciences et des techniques
M. LAURE LUCOT	CLINIQUE URM 7137 - Ecole INSPI Fondation Makovici	66	Réhabilitation
Mme WALLEBS-OR	Département d'Histologie, Anatomie Médicale, Faculté de Médecine	72	Endocrinologie - Histoire des sciences et des techniques
Mme GILBERT-MERCIER	Département d'Histologie, Anatomie Médicale, Faculté de Médecine	72	Cardiologie - Histoire des sciences et des techniques
Mme T. COVAS-MORIN	Département d'Histologie, Anatomie Médicale, Faculté de Médecine	72	Cardiologie - Histoire des sciences et des techniques
M. VIGNON-BOUILLON	Département d'Histologie, Anatomie Médicale, Faculté de Médecine	72	La cardiologie - Histoire des sciences et des techniques
M. ZIVICH-AYOUB	Département d'Histologie, Anatomie Médicale, Faculté de Médecine	72	Endocrinologie - Histoire des sciences et des techniques

C - ENSEIGNANTS ASSOCIES DE MEDECINE GENERALE

C1 - PROFESSEURS ASSOCIES DES UNIVERSITES DE M. G. (mi-temps)

Pr. Ass. GREGOIRE-JOUEK	4004	Médecine générale (01/09/2017)
M. Ass. GREGOIRE-JOUEK	4072	Médecine générale (01/09/2015)
Pr. Ass. GREGOIRE-FRAPPY	4005	Médecine générale (01/11/2015)
Pr. Ass. FLOU-THARON	4008	Médecine générale (01/11/2015)
Yves HOUSSIER-THARON	4007	Médecine générale (01/09/2015)

C2 - MAITRE DE CONFERENCES DES UNIVERSITES DE MEDECINE GENERALE - TITULAIRE

Dr. GABRIEL-LUCOT	4015	Médecine générale (01/09/2010)
Dr. SCHMIDT-THARON		Médecine générale

C3 - MAITRES DE CONFERENCES ASSOCIES DES UNIVERSITES DE M. G. (mi-temps)

Dr. GREGOIRE-JOUEK	Médecine générale (01/09/2015) / (01/09/2016)
Dr. GABRIEL-LUCOT	Médecine générale
Dr. SCHMIDT-THARON	Médecine générale

D - ENSEIGNANTS DE LANGUES ETRANGERES

D1 - PROFESSEUR AGREGÉ, PRAG et PRCE DE LANGUES

Mme ACHILLE-SCOTTI	4006	Professeure agrégée d'Espagnol (depuis 01/09/10)
Mme CHIFFOLETTI	4004	Professeure agrégée d'Espagnol (depuis le 01/09/10)
Mme BENEDEK-THARON	4004	Professeure agrégée d'Espagnol (depuis le 01/09/10)
Mme J. BARBERIS	4008	Professeure agrégée d'Espagnol (depuis 01/09/10)
Mme MARTIN-BOURNE	4006	Professeure agrégée d'Espagnol (depuis 01/09/10)

Serment d'Hippocrate

« Au moment d'être admis 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 dans l'intimité des personnes, je tairai les secrets qui me sont confiés. Reçu 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 et méprisé si j'y manque ».

Dédicaces et Remerciements

À mon président de thèse, le Professeur PELACCIA Thierry, qui me fait l'honneur de présider ce jury de thèse. Je vous remercie de m'avoir permis de réaliser ce travail.

À mon directeur de thèse, le Docteur FABACHER Thibaut, que je remercie pour sa disponibilité pendant la réflexion, la conception de l'algorithme, l'écriture de cette thèse et pour ses conseils. Ce fut un plaisir de travailler avec toi.

Aux membres du jury, Docteure CHAMBE Juliette et Docteur LECOINTRE Lise, je vous prie de recevoir mes sincères remerciements pour avoir accepté de juger mon travail. Vous me faites l'honneur d'apporter votre expérience à ce travail en siégeant dans ce jury.

Ensuite je souhaite remercier les professeur(e)s de la faculté de médecine, pour leurs enseignements et leurs appuis, tout au long de notre formation.

Je remercie les professionnels ayant pris de leur temps pour m'avoir permis d'entamer une réflexion autour de ce sujet et pour avoir contribué à la construction de mon identité professionnelle.

Je remercie mes amis et mes collègues pour leur aide, leur soutien et leur investissement.

A Aurélien qui m'a inspiré pour cette thèse, sur qui j'ai pu toujours compter, et qui m'a aidé en tant que relecteur.

A Nella, Efflam, Raphaëlle, Manon, Floriane, Zsölt, Elise, Mégane et Juliette sans qui mes années d'externat auraient été beaucoup plus difficiles.

A Pilou, Serviette, Cyril, Nico, Kevin, Cyrielle, Émilie et Rémi pour m'avoir permis de décrocher un peu du boulot.

Merci à Virginie et Marie Laurence, pour avoir pris sur leur temps pour me relire.

Je remercie ma famille, et plus particulièrement ma sœur et mes parents pour leur soutien infailible, leur aide au quotidien et leur patience.

Je remercie également mes beaux-parents pour leur présence, et leur aide.

A mon chat, Podrick, qui n'a pas fait grand-chose à part la sieste.

Enfin, j'ai une pensée pour ma concubine Orane qui m'a soutenu dans ces années difficiles avec toujours autant de patience et de compréhension. Et avec qui j'espère pouvoir finir ma vie.

À toutes ces personnes, et celles que j'ai omis, mes pensées les plus sincères pour tout ce qu'elles m'ont apporté.

I. Table des matières

Serment d’Hippocrate	12
Dédicaces et Remerciements	13
Glossaire	21
Liste des annexes.....	25
Liste des tableaux, figures et équations.....	27
1 Introduction.....	30
1.1 Historique des méthodes de fouilles de textes	31
1.1.1 La loi de Moore.....	31
1.2 Historique de la médecine générale	32
1.3 Évolutions récentes et définitions de la médecine générale.....	33
1.4 Évolution, état des lieux et enjeux de la recherche en médecine générale.....	35
1.5 Le troisième cycle des études médicale en médecine générale et la thèse.....	37
2 Méthode et matériel	40
2.1 Méthode	40
2.1.1 Type d’étude.....	40
2.1.2 Extraction des données	40
2.1.3 Tri des données	41
2.1.3.1 Séparation des fiches de thèses entre elles	41
2.1.3.2 Recherche et tri des données présentes dans chaque fiche.....	42
2.1.3.3 Harmonisation de certains lieux d’origine des fiches de thèse.....	43
2.1.3.4 Exclusion des fiches de thèses ne contenant pas de résumés.....	44
2.1.4 Exploitation des données globales.....	44

2.1.4.1	Estimation de l'âge des auteurs et de l'âge moyen de passage de la thèse	44
2.1.4.2	Répartition géographique de l'échantillon.....	44
2.1.4.3	Effectif et proportion des thèses analysées par années	44
2.1.5	Préparation du corpus en vue de son analyse	45
2.1.5.1	Mise en forme de la base de données	45
2.1.5.2	Lemmatisation.....	45
2.1.5.3	Racinisation	45
2.1.5.4	Suppression des mots vides ou stopwords	45
2.1.6	Occurrences totales des mots	46
2.1.7	Décompte des paires de mots au sein d'un même document.....	46
2.1.8	Corrélation entre les différents mots.....	46
2.1.9	La fréquence de terme et la fréquence de document inverse.....	47
2.1.10	Analyse en bigramme	48
2.1.11	Approche par apprentissage automatique.	48
2.1.11.1	Échantillonnage de Gibbs	49
2.1.11.2	Allocation de Dirichlet latente.....	49
2.1.12	Évolution des sujets de thèses de 2015 à 2021.....	50
2.1.12.1	Analyse par fréquence de terme dans le corpus.....	50
2.1.12.2	Analyse et comparaison de l'évolution par LDA.....	50
2.2	Matériel	51
3	Résultats.....	52
3.1	Données globales de l'échantillon	52
3.1.1	Taille de l'échantillon et période d'étude retenue.....	52
3.1.2	Répartition géographique de l'échantillon.....	52
3.1.3	Âge des auteurs lors de la soutenance de la thèse	54
3.1.3.1	Âge des auteurs lors de la soutenance de la thèse	55
3.1.4	Effectif des fiches de thèses par années	56

3.2	Analyse par fouille de texte	58
3.2.1	Décompte de tous les termes	58
3.2.1.1	Décompte de tous les termes toutes années confondues.....	58
3.2.1.2	Décompte de tous les termes par année	59
3.2.1.3	Décompte des termes sur l'année 2021	61
3.2.2	Fréquence d'association des mots entre eux en 2021.....	62
3.2.3	Corrélation des termes entre eux	63
3.2.3.1	Corrélation des termes entre eux sans tri des termes	63
3.2.3.2	Corrélation des termes entre eux avec tri des termes.....	64
3.2.3.2.1	Détail des corrélations des termes entre eux avec tri des termes.....	65
3.2.4	Fréquence de terme et fréquence de document inverse	69
3.2.5	Analyse des bigrammes.....	71
3.2.5.1	Décompte de tous les bigrammes les plus fréquents par année	71
3.2.5.2	Fréquence d'association des bigrammes entre eux en 2021.....	72
3.2.5.3	Corrélation des bigrammes entre eux avec tri des termes	73
3.2.5.4	Fréquence des bigrammes et fréquence de document inverse.....	74
3.3	Fouille de texte par apprentissage automatique	75
3.3.1	Échantillonnage de Gibbs sur l'année 2021	75
3.3.2	Allocation latente de Dirichlet.....	77
3.3.2.1	Résultats de l'Allocation latente de Dirichlet	77
3.3.2.2	Identification de sujets via les résultats obtenus	79
3.4	Évolution de certains sujets depuis 2015.....	80
3.4.1	Résultats avec augmentation significative	81
3.4.1.1	Télé médecine	81
3.4.1.2	La relation aux écrans.....	83
3.4.1.3	La COVID	85
3.4.1.4	Violences	87
3.4.1.5	Le parcours de soin.....	89
3.4.1.6	L'aspect Social	91

3.4.1.7	Vaccination	93
3.4.1.8	Activité sportive.....	95
3.4.1.9	Allaitement	97
3.4.1.10	Suivi pédiatrique en médecine générale.....	99
3.4.1.11	Suivi gynécologique et contraception chez la femme.....	102
3.4.1.12	Installation et exercice de la médecine générale en milieu rural	104
3.4.1.13	La prévention en médecine générale.....	106
3.4.2	Résultats avec diminution significative	108
3.4.2.1	Hospitalisation et personnes âgées.....	108
3.4.2.2	Soins à domicile	111
3.4.3	Résultats avec augmentation non significative	113
3.4.4	Résultats avec diminution non significative	114
3.5	Résultats par comparaison directe en LDA	115
4	Discussion.....	117
4.1	Discussion des principaux résultats	117
4.1.1	Téléconsultation.....	117
4.1.2	Écrans	118
4.1.3	COVID	119
4.1.4	Violence.....	119
4.1.5	Parcours.....	120
4.1.6	Social.....	121
4.1.7	Vaccination.....	121
4.1.8	Sport	123
4.1.9	Allaitement	123
4.1.10	Parentalité et enfant	124
4.1.11	Contraception.....	124

4.1.12	Prévention	125
4.1.13	Installation	125
4.1.14	Santé mentale	126
4.1.15	Hospitalisation	127
4.1.16	Personnes âgées	127
4.1.17	Domicile	128
4.1.18	Diabète	129
4.1.19	Anticoagulants	129
4.2	Validité interne	130
4.2.1	Avantages de cette étude	130
4.2.1.1	Taille de l'échantillon	130
4.2.1.2	Extensibilité	130
4.2.1.3	Analyse ciblée	130
4.2.1.4	Multicentrique	130
4.2.1.5	Reproductibilité	132
4.2.1	Désavantages de cette étude	132
4.2.1.1	Biais	132
4.2.1.1.1	Biais de sélection	132
4.2.1.1.2	Biais de classement	132
4.2.1.1.3	Biais d'évaluation	133
4.2.1.1.4	Biais de mesure	133
4.2.1.1.5	Biais de représentativité	133
4.2.1.2	Complexité	133
4.2.1.2.1	Choix des mots vides de sens	133
4.2.1.2.2	Choix des termes utilisées pour la recherche dans la base de données	134
4.3	Validité externe	134
4.3.1	Pertinence du choix de la base de données	134

4.3.2	Période d'étude	134
4.3.3	Origine géographique des thèses	135
4.3.4	Effectifs	135
4.3.5	Tri des sujets appartenant à la médecine générale ou non	135
4.3.6	Taxonomie des sujets	136
4.3.7	Évolution des thématiques.....	136
4.3.8	Comparaison des résultats	137
4.3.8.1	Âge moyen.....	137
4.3.8.2	Sujets prépondérants	137
4.3.8.3	Évolution des sujets.....	140
5	Conclusion	142
6	Annexes	143
7	Bibliographie.....	175

Glossaire

ADL : Activities of Daily Living (échelle permettant de quantifier l'autonomie d'une personne)

Amelipro : site internet qui accueille des téléservices pour les professionnels de santé

BITS-test : Brimade Insomnie Tabac Stress (repérage des problématiques suicidaires et des scarifications des 13-18 ans)

CCI : Charlson Comorbidity Index (L'indice de comorbidité de Charlson)

CeGIDD : Centres gratuits d'information, de dépistage et de diagnostic

CERDAM : Centre de Ressources Documentaires sur l'Allaitement Maternel (1)

CHU : centre hospitalo-universitaire

COVID : Corona virus disease

CNGE : Collège national des généralistes enseignants

CMVV : patients cardio-vasculaires et métaboliques

DES : diplôme d'études spécialisées

DIU : dispositif intra-utérin

EDC : Épisode dépressif caractérisé

EGS : Évaluation gériatrique standardisée

EHPAD : établissement d'hébergement pour personnes âgées dépendantes

EMSP : équipe mobile de soins palliatifs

ETHOS : European Typology on Homelessness and housing exclusion (Typologie européenne de l'exclusion liée au logement)

EVA : Échelle visuelle analogique (mesure l'intensité de la douleur sur une échelle allant de 0 à 10)

Eval Mater : bilan de santé Eval Mater de la Protection Maternelle et Infantile

EXDEMPAGE : Etude EXDEMPAGE : enquête sur la santé et les habitudes de vie de collégiens de la Métropole Européenne de Lille avant une intervention ciblée (2)

FDR : Facteurs de risque

GIR : Groupe iso-ressources (niveau de perte d'autonomie d'une personne âgée)

HAD : Hospital Anxiety and Depression scale

HbA1c : hémoglobine glyquée

HPV : Human Papillomavirus

HSH : Homme ayant des rapports sexuels avec un ou d'autres hommes

IAH : Indice d'Apnée Hypopnée

IOTF-30 : International Obesity Task Force : une définition du surpoids et de l'obésité chez l'enfant (3)

IRI : Interpersonal Reactivity Index

IST : Infection sexuellement transmissible

ITT : Incapacité Temporaire de Travail

IVG : interruption volontaire de grossesse

SFMG : Société française de médecine générale

Lactaclic : Site d'informations pratiques sur l'allaitement maternel en soins primaires (4)

LDA : Allocation latente de Dirichlet

MBI : Maslach burnout inventory ,questionnaire d'évaluation de l'épuisement professionnel (5)

MBSR : Mindfulness Based Stress Reduction (réduction du stress basée sur la pleine conscience)

MedByColisée : Un logiciel de gestion médical, prescription et téléconsultation

MEL : Métropole européenne de Lille

MG : Médecin généraliste/Médecine générale

MPC : méditation en pleine conscience

NHC : Nouvel hôpital civil

OMS : Organisation mondiale de la santé

PCLS : Post-traumatic stress disorder Checklist Scale (Echelle de l'état de stress post-traumatique)

PECARN : Pediatric Emergency Care Applied Research Network

PhysiCOVID : étude sur l'évolution des modalités de pratique d'une activité physique et sportive, de l'anxiété et du sommeil dans les suites de la période de confinement imposé par la pandémie de COVID-19 chez des patients porteurs d'une pathologie chronique et chez des sujets sains sportifs.

PMI : Protection Maternelle et Infantile

PSS : Perceived Stress Scale (Echelle de stress perçu.)

PTA : plateforme territoriale d'appui

PTSD : Post-traumatic stress disorder (troubles du stress post-traumatique)

RÉPPOP : Réseau de Prévention et de Prise en charge de l'Obésité Pédiatrique

SAHOS : syndrome d'apnées hypopnées obstructives du sommeil

SAOS : syndrome d'apnées obstructives du sommeil

SAU : Service d'accueil des urgences

SECCA : Situation Emotions Cognitions Comportements Analyse

STAI : State Trait Anxiety Inventory

TF-IDF : Terme Frequency et Inverse Document Frequency

TLC : téléconsultation

TLCA : téléconsultation assistée

TCG : Cotation d'une téléconsultation standard en médecine générale

TSN : traitement substitutif nicotinique

TSPT : troubles du stress post-traumatique

TSTS-CAFARD : acronyme d'une série de questions pour évaluer le mal-être chez l'adolescent.

UMJ : Unité médico-judiciaire

USP : Unités de Soins Palliatifs

VIH : virus de l'immunodéficience humaine

VSRQ : Visual Simplified Respiratory Questionnaire (Questionnaire respiratoire visuel simplifié)

Liste des annexes

Annexe 1 : Liste des termes changés pour la lemmatisation du corpus	143
Annexe 2 : Liste des mots vides de sens appliqués à l'ensemble de l'analyse.....	145
Annexe 3:Liste des mots supprimés de l'analyse de corrélation entre les termes	147
Annexe 4 : Liste des mots supprimés de l'analyse de corrélation entre les bigrammes	148
Annexe 5: Essai sur les maladies des artisans page 374	149
Annexe 6 : Évolution et corrélations de l'expression infections sexuellement transmissibles sur la période 2015 à 2021.....	150
Annexe 7 : Évolution et corrélations du terme maltraitance sur la période 2015 à 2021	151
Annexe 8 : Évolution et corrélations du terme anxiété sur la période 2015 à 2021	152
Annexe 9 : Évolution et corrélations du terme dépression sur la période 2015 à 2021.....	153
Annexe 10 : Évolution et corrélations du terme stress sur la période 2015 à 2021	154
Annexe 11 : Évolution et corrélations de l'expression épuisement professionnel sur la période 2015 à 2021.....	155
Annexe 12 : Évolution et corrélations du terme sommeil sur la période 2015 à 2021.....	156
Annexe 13 : Évolution et corrélations du terme travail sur la période 2015 à 2021	157
Annexe 14 : Évolution et corrélations du terme adolescent sur la période 2015 à 2021.....	158
Annexe 15 : Évolution et corrélations du terme surpoids sur la période 2015 à 2021.....	159
Annexe 16 : Évolution et corrélations du terme obésité sur la période 2015 à 2021	160
Annexe 17 : Évolution et corrélations du terme femme sur la période 2015 à 2021.....	161
Annexe 18: Évolution et corrélations du terme formation sur la période 2015 à 2021	162
Annexe 19 : Évolution et corrélations du terme pédiatrie sur la période 2015 à 2021.....	163
Annexe 20 : Évolution et corrélations du terme tabagisme sur la période 2015 à 2021.....	164
Annexe 21 : Évolution et corrélations du terme urgence sur la période 2015 à 2021	165
Annexe 22 : Évolution et corrélations du terme gériatrie sur la période 2015 à 2021.....	166

Annexe 23 : Évolution et corrélations du terme diabète sur la période 2015 à 2021	167
Annexe 24 : Évolution et corrélations du terme prescription sur la période 2015 à 2021	168
Annexe 25 : Évolution et corrélations du terme palliatif sur la période 2015 à 2021	169
Annexe 26 : Évolution et corrélations du terme homme sur la période 2015 à 2021	170
Annexe 27 : Évolution et corrélations du terme sevrage sur la période 2015 à 2021	171
Annexe 28 : Évolution et corrélations du terme alcool sur la période 2015 à 2021	172
Annexe 29 : Évolution et corrélations du terme grossesse sur la période 2015 à 2021	173
Annexe 30 : Évolution et corrélations du terme douleur sur la période 2015 à 2021	174

Liste des tableaux, figures et équations

Tableau 1: Liste des noms de villes remplacés.....	43
Tableau 2 : Répartition géographique des fiches de thèse récupérées.....	53
Tableau 3 : Probabilités de présence d'un sujet dans un résumé de thèse obtenue par allocation latente de Dirichlet et interprétations.....	77
Tableau 4: Résultats de la LDA sur la période 2015 à 2021, avec l'évolution moyenne par année (* $p < 0,05$)	115
Figure 1 : Schéma représentant le fonctionnement du programme de tri des données	42
Figure 2 : Répartition géographique du nombre de fiches de thèses analysées	52
Figure 3 : Âge des auteurs lors de la soutenance de thèse.....	54
Figure 4 : Âge moyen des auteurs lors de la soutenance de thèse.....	55
Figure 5 : Représentation graphique du nombre de fiches de thèses récupérées par années	56
Figure 6 : Nombre de termes par résumé.....	57
Figure 7 : Représentation des mots les plus fréquents par nuage de mots.....	58
Figure 8 : Nombre d'occurrences des 10 termes les plus fréquents par année	59
Figure 9 : Présence des termes "patient" et "médecin" dans chaque résumé de thèse	60
Figure 10 : Les 20 termes les plus cités dans les résumés de thèses en 2021	61
Figure 11 : Fréquence d'association des termes entre eux.....	62
Figure 12 : Corrélations de plus de 50% entre les termes apparaissant au moins 20 fois dans les thèses de 2021 sans tri	63
Figure 13 : Corrélations de plus de 25% entre les termes apparaissant au moins 20 fois dans les thèses de 2021 avec tri.....	64
Figure 14 : Agrandissement du quadrant inférieur droit de la Figure 13.....	65
Figure 15 : Agrandissement du quadrant supérieur gauche de la Figure 13	66

Figure 16 : Agrandissement du quadrant inférieur gauche de la Figure 13.....	67
Figure 17 : Agrandissement du quadrant inférieur de la Figure 13	68
Figure 18 : Les 20 termes ayant le TF-IDF le plus grand en 2021	69
Figure 19 : Graphique du TF-IDF en fonction du nombre d'occurrences en 2021	70
Figure 20 : Décompte des bigrammes les plus fréquents de 2015 à 2021	71
Figure 21 : Fréquence d'association des bigrammes entre eux sans tri des termes.....	72
Figure 22 : Corrélations de plus de 25 % des bigrammes apparaissant plus de 10 fois dans les résumés de thèses en 2021 avec tri des termes.....	73
Figure 23 : Nuage de mots des 10 plus grands TF-IDF des bigrammes de 2021	74
Figure 24 : Échantillonnage de Gibbs de 10 à 100 sujets en 2021	75
Figure 25 : Détermination du sujet de deux résumés de thèses du corpus.....	79
Figure 26 : Évolution et corrélations du terme téléconsultation sur la période 2016 à 2021	81
Figure 27 : Évolution et corrélations du terme écran sur la période 2015 à 2021	83
Figure 28 : Évolution et corrélations du terme COVID sur la période 2020 à 2021	85
Figure 29 : Évolution et corrélations du terme violence sur la période 2015 à 2021	87
Figure 30 : Évolution et corrélations du terme parcours sur la période 2015 à 2021	89
Figure 31 : Évolution et corrélations du terme social sur la période 2015 à 2021	91
Figure 32 : Évolution et corrélations du terme vaccination sur la période 2015 à 2021	93
Figure 33 : Évolution et corrélations du terme sport sur la période 2015 à 2021	95
Figure 34 : Évolution et corrélations du terme allaitement sur la période 2015 à 2021	97
Figure 35 : Évolution et corrélations du terme parentalité sur la période 2015 à 2021.....	99
Figure 36 : Évolution et corrélations du terme enfant sur la période 2015 à 2021	100
Figure 37 : Évolution et corrélations du terme contraception sur la période 2015 à 2021	102
Figure 38 : Évolution et corrélations du terme installation sur la période 2015 à 2021.....	104
Figure 39 : Évolution et corrélations du terme prévention sur la période 2015 à 2021.....	106
Figure 40 : Évolution et corrélations du terme hospitalisation sur la période 2015 à 2021	108

Figure 41 : Évolution et corrélations du terme âgé sur la période 2015 à 2021.....	109
Figure 42 : Évolution et corrélations du terme domicile sur la période 2015 à 2021.....	111
Figure 43 : Nombre de visites à domicile en France de 1993 à 2013, d'après les données d'éco-santé (70)	112
Équation 1 : Corrélation entre les termes.....	47
Équation 2 : Calcul du TF	47
Équation 3 : Calcul de l'IDF	47

1 Introduction

Durant mon troisième cycle de diplôme d'études spécialisées (DES) de médecine générale, je cherchais un sujet de thèse. Malheureusement mes projets de sujets de thèse n'ont pas abouti. Lors du début de ma deuxième année de post internat, alors que j'étais médecin remplaçant, j'ai pu assister à la thèse d'un de mes amis et confrère.

Cela me donna l'idée de chercher un sujet de manière plus méthodique, avec l'appui du guide pratique du thésard (6). Après avoir dressé une liste de tous les sujets qui seraient susceptibles de m'intéresser, je suis allé consulter les bases de données de thèses de médecine générale, en particulier la base de données Sudoc (7). Les sujets abordés étaient assez proches de ceux qui pouvaient m'intéresser, et semblaient assez fréquents.

En utilisant ma liste comme modèle je me suis donc mis à compter sur plusieurs pages de résultats les thèses qui correspondaient aux sujets que j'avais présélectionnés. Il y avait effectivement des thèmes qui étaient très populaires, ce qui me donna envie de savoir quels sujets intéressaient tant les thésards de médecine générale.

Après quelques semaines de recherche, j'ai pu trouver mon directeur de thèse et de mémoire. À la suite de quelques essais de classification des thèmes sur des petits effectifs, nous avons pu effectivement observer des redondances dans les sujets abordés et ainsi commencer à travailler sur l'ensemble des effectifs disponibles de l'année 2020 pour mon mémoire « Analyse des sujets de thèses de médecine générale en France en 2020 » soutenu en 2021 puis des années 2015 à 2021 pour cette thèse.

Différents travaux ont déjà exploré les sujets, la méthodologie et la qualité des thèses de médecine générale (8–11). Notre approche va plutôt se concentrer sur une analyse via la fouille de texte ou text mining. Cette approche pourrait permettre de trouver des sujets de manière plus précise,

de travailler sur des effectifs plus grands et d'offrir, en plus, une plus grande reproductibilité. Cette méthode a déjà été utilisée en 2016 afin d'analyser des archives médicales du XIX^{ème} au XX^{ème} siècle (12).

1.1 Historique des méthodes de fouilles de textes

Les méthodes de fouilles de textes ont évolué au fil du temps, avec l'avancée de la technologie informatique. Cependant, les limitations technologiques en informatiques et le manque de données ont fait que les recherches sur le sujet sont très limitées ou théoriques avant les années 1990.

C'est au cours des années 1990 et 2000, que les méthodes de fouilles de textes ont continué à s'améliorer avec l'avènement des technologies de l'Internet et des réseaux de données(13,14). Les outils de fouille de textes ont été développés pour traiter de grandes quantités de données issues de sources numériques telles que les journaux en ligne, les réseaux sociaux et les archives électroniques. Les méthodes de traitement automatique du langage naturel ont été développées pour extraire des informations à partir de textes non structurés, comme les messages de forums et les commentaires sur les blogs.

Actuellement, les méthodes de fouille de textes continuent à évoluer avec les avancées dans les domaines de l'apprentissage automatique et de l'intelligence artificielle, permettant de traiter de plus en plus de données complexe, d'automatiser les tâches de traitement, d'augmenter la précision et de créer des visualisations plus élaborées. Les fouilles de textes sont de plus en plus utilisées pour analyser des sources numériques, comme les tweets, les articles de blog et les données de réseaux sociaux.

1.1.1 La loi de Moore

La loi de Moore est une loi empirique en informatique qui prévoit que le nombre de transistors intégrés sur une puce double environ tous les 18 à 24 mois. Cette loi a été formulée pour la première fois par Gordon Moore, co-fondateur d'Intel, dans un article publié en 1965(15). Il a observé que la

densité de transistors sur les puces électroniques augmentait de manière exponentielle, et il a prédit que cette tendance se poursuivrait à l'avenir.

La loi de Moore a été un moteur clé de la croissance de l'industrie de l'informatique au cours des dernières décennies. Les progrès réalisés dans la technologie des transistors ont permis une augmentation constante de la puissance de calcul des ordinateurs, ce qui a entraîné l'apparition de nouvelles applications informatiques, comme la robotique, la vision par ordinateur, l'apprentissage automatique et l'intelligence artificielle.

Les méthodes de fouilles de textes présente donc aussi l'avantage de bénéficier des avancées en matière informatique de manière régulières, permettant de reproduire des recherches avec plus de précisions et de rapidité d'année en année.

1.2 Historique de la médecine générale

La médecine générale en France a évolué au cours des siècles. Au Moyen Âge, les médecins étaient généralement des moines ou des religieux qui pratiquaient la médecine en plus de leurs autres tâches religieuses(16). Au cours des siècles suivants, la médecine s'est professionnalisée, et les médecins ont commencé à étudier et à pratiquer la médecine de manière plus scientifique.

Au 19^{ème} siècle, la médecine générale en France a été transformée par l'émergence de nouvelles technologies et de nouveaux traitements. L'utilisation de la stérilisation, de l'anesthésie et de la radiologie(17) ont considérablement amélioré les soins médicaux. La médecine générale a également été influencée par les développements en matière d'hygiène, de nutrition et de vaccination.

Au cours des 20^{ème} siècles, la médecine générale en France a continué à évoluer avec l'arrivée des nouvelles technologies et des avancées médicales. Les médecins généralistes sont devenus les médecins de première ligne et ont été chargés de la coordination des soins pour les patients atteints de maladies chroniques. Les études médicales ont été remaniées pour inclure des stages de formation en médecine générale(18) dans le cursus.

Aujourd'hui, la médecine générale en France est un domaine complexe, qui traite un grand nombre de pathologies différentes. Les médecins généralistes jouent un rôle important dans la prévention des maladies, ainsi que dans la prise en charge des patients atteints de maladies chroniques. Ils jouent également un rôle important dans la coordination des soins et dans la communication avec les autres professionnels de la santé.

1.3 Évolutions récentes et définitions de la médecine générale en France

La médecine générale a été promue ces dernières années en France, notamment grâce à la création de postes de généralistes enseignants en 1983(19,20), l'obligation de faire un stage chez les généralistes pendant le troisième cycle en 1997(18,20), la création du diplôme d'études spécialisées de médecine générale en 2004(20,21), et la titularisation de professeurs de médecine générale en 2009(20,22).

En 1974, le groupe de Leeuwenhorst a défini la médecine générale comme les soins prodigués par un médecin généraliste pour les personnes, les familles et la population, indépendamment de l'âge, du sexe et de la maladie(23). Cette définition est surtout basée sur les activités professionnelles, mais ne prend pas en compte les principes fondamentaux en tant que discipline scientifique. En 1978, l'OMS a défini les soins primaires comme étant des soins de santé essentiels accessibles à tous par des moyens acceptables, avec la participation des patients et à un coût abordable. En 1972, la WONCA (World Organization of National Colleges, Academies and Academic Associations of General Practitioners) a été créée pour promouvoir la médecine générale et développer cette discipline à l'échelle mondiale. La définition actuelle de la médecine générale est celle de WONCA Europe de 2002 qui précise que c'est une discipline scientifique et universitaire qui se consacre aux soins primaires(23).

La définition de la WONCA Europe de 2002 de la médecine générale précise également que c'est une spécialité clinique qui s'occupe des soins primaires et qui possède son propre contenu de

formation, de recherche, de pratique clinique et des fondements scientifiques. La médecine générale est considérée comme une spécialité médicale à part entière, avec la recherche comme une partie intégrante.

Les 11 caractéristiques qui décrivent cette spécialité sont :

- *« Elle est habituellement le premier contact avec le système de soins, permettant un accès ouvert et non limité aux usagers, prenant en compte tous les problèmes de santé, indépendamment de l'âge, du sexe, ou de toutes autres caractéristiques de la personne concernée.*
- *Elle utilise de façon efficiente les ressources du système de santé par la coordination des soins, le travail avec les autres professionnels de soins primaires et la gestion du recours aux autres spécialités, se plaçant si nécessaire en défenseur du patient.*
- *Elle développe une approche centrée sur la personne dans ses dimensions individuelle, familiale, et communautaire.*
- *Elle utilise un mode de consultation spécifique qui construit dans la durée une relation médecin-patient basée sur une communication appropriée.*
- *Elle a la responsabilité d'assurer des soins continus et longitudinaux, selon les besoins du patient.*
- *Elle base sa démarche décisionnelle spécifique sur la prévalence et l'incidence des maladies en soins primaires.*
- *Elle gère simultanément les problèmes de santé aigus et chroniques de chaque patient.*
- *Elle intervient à un stade précoce et indifférencié du développement des maladies, qui pourraient éventuellement requérir une intervention rapide.*
- *Elle favorise la promotion et l'éducation pour la santé par une intervention appropriée et efficace.*
- *Elle a une responsabilité spécifique de santé publique dans la communauté.*

- *Elle répond aux problèmes de santé dans leurs dimensions physique, psychologique, sociale, culturelle et existentielle. »(23).*

Cependant, en France, cette spécialité n'a été reconnue qu'à partir de 2004 avec l'introduction du Diplôme d'Etudes Spécialisées de médecine générale.

1.4 Évolution, état des lieux et enjeux de la recherche en médecine générale

La médecine générale est un champ de compétences large qui comprend la prévention, le dépistage, la surveillance, l'orientation et l'éducation en plus de ses fonctions diagnostiques et thérapeutiques. La recherche en médecine générale est nécessaire pour améliorer la qualité des soins primaires en améliorant la compréhension et la pratique. Cependant, jusqu'à la moitié du XXème siècle, la médecine générale n'était pas encore considérée comme une spécialité en soi, et la France a développé un retard par rapport à ses voisins européens en matière de recherche en médecine générale. (24).

La recherche en médecine générale est importante pour améliorer la qualité des soins primaires et de la pratique médicale en général. Il existe de nombreux travaux de recherche en médecine hospitalière qui peuvent fournir des informations utiles pour la médecine ambulatoire, cependant, l'utilisation des données et des résultats obtenus dans la recherche hospitalière pour la médecine de ville est discutable. En effet, les pathologies et les populations de patients entre les deux milieux sont différents. Les critères d'inclusion utilisés dans les études cliniques hospitalières ne sont pas toujours applicables à la médecine générale. En outre, certains aspects importants de l'exercice de la médecine générale, tels que la prévention, le dépistage, l'éducation et les relations médecin-patient, ne peuvent être étudiés adéquatement dans un contexte hospitalier.

Les travaux de Kerr White, F. Williams, et B. Greenberg ont montré en 1961 que les patients qui consultent en CHU ne représentent qu'une très faible proportion de la population exposée à un

problème de santé chaque mois, de l'ordre de 1 pour 1000(25). Ces résultats semblent stables dans le temps, d'après une autre étude parue en 2001 (26). Ils ont également mis en évidence l'importance des soins de première ligne ou soins primaires, ainsi que du rôle du médecin généraliste en tant que coordinateur des soins. Les résultats de ces études ont conduit à une réflexion sur la distribution des soins en population générale et ont contribué au développement de la médecine générale et de la recherche en soins primaires.

Les publications sur les soins primaires en France sont moins nombreuses que dans certains autres pays(27), malgré l'importance de cette recherche pour améliorer la pratique médicale. Il y a plusieurs barrières qui dissuadent les médecins généralistes français de se lancer dans la recherche, notamment le manque d'opportunités et les difficultés à obtenir les fonds nécessaires(28).

Les départements de médecine générale (DMG) sont des structures universitaires qui organisent les études de médecine générale. Ils ont pour mission de concevoir, de mettre en œuvre et d'évaluer les programmes de formation pour les étudiants en médecine ainsi que pour les médecins en exercice. Les DMG ont également pour rôle de favoriser la recherche en médecine générale et d'encourager les médecins généralistes à se perfectionner tout au long de leur carrière. Les DMG sont généralement rattachés aux facultés de médecine et de pharmacie des universités françaises et ils travaillent en étroite collaboration avec les syndicats médicaux et les organisations professionnelles de médecins généralistes(29).

Des initiatives qui ont été mises en place pour promouvoir la recherche en médecine générale en France. Le collège national des généralistes enseignants (CNGE), la société française de médecine générale (SFMG), la société française de documentation et recherche en médecine générale (SFDRMG), la société française de thérapeutique généraliste (SFTG), les départements de médecine générale et les réseaux locaux ont des groupes de travail ou des programmes spécifiques pour soutenir les projets de recherche menés par les généralistes enseignants et les encourageant à poursuivre dans cette voie(30–33). Les universités proposant des masters avec des enseignements de recherche

peuvent aider les étudiants à développer les compétences nécessaires pour devenir des chercheurs en médecine générale. Enfin, les instituts publics de recherche comme l'institut national de la santé et de la recherche médicale (INSERM) et l'agence nationale de sécurité du médicament et des produits de santé (ANSM) ont des appels à projets ouverts aux médecins généralistes pour leur permettre de soumettre leurs idées de recherche et d'obtenir des financements pour les mener à bien(34,35).

La recherche en médecine générale est cruciale pour améliorer la qualité des soins primaires et pour répondre aux besoins de santé spécifiques des patients en soins primaires. Les thèses de médecine générale couvrent un large éventail de sujets, allant de la prévention, au dépistage, à la surveillance, à l'orientation et à l'éducation en plus des fonctions diagnostique et thérapeutique. Il est important de décrire les thèses soutenues en médecine générale pour identifier les thématiques dominantes.

1.5 Le troisième cycle des études médicale en médecine générale et la thèse

Le décret n° 2004-67 du 16 janvier 2004 énonce les règles et les procédures pour l'accès, l'organisation des stages et des enseignements, ainsi que les conditions d'obtention du diplôme d'Etat de docteur en médecine générale pour les études médicales de troisième cycle (TCEM) en France(36). Une réforme du TCEM a été mise en place à la rentrée 2017 qui visait à moderniser le système de formation des médecins, et à simplifier les règles en place(24,37). Pour les internes "ancien régime", c'est à dire ceux avant la réforme, les outils de validation du Diplôme d'Etat de médecine générale (DES) comprennent une formation de 3 ans, un programme d'enseignements obligatoires qui comprend 4 semestres dans différents services, deux stages libres, une thèse, dont le sujet n'est pas forcément en relation directe avec la médecine générale et un mémoire de DES.

La réforme de 2017 du Diplôme d'Etat de médecine générale (DES) a apporté des changements importants à la structure de la formation et à l'évaluation des compétences des internes en médecine générale en France. Les objectifs de cette réforme étaient de moderniser le système de formation des

médecins généralistes, de simplifier les règles existantes, de permettre une progression continue des compétences professionnelles et d'améliorer la reconnaissance des qualifications. Les compétences à acquérir incluent des compétences génériques et transversales liées à la démarche décisionnelle du premier recours ambulatoire, à l'approche centrée patient, et à l'approche globale, ainsi que des compétences spécifiques liées aux 6 compétences génériques du médecin généraliste(38). La formation est d'une durée de 6 semestres et il y a des évaluations périodiques pour s'assurer que les objectifs pédagogiques sont atteints et que les capacités de formation sont développées(37,39).

La réforme du TCEM de 2017 a modifié les exigences en matière de production scientifique pour les internes en médecine générale. En particulier, elle a remplacé le mémoire de DES par un portfolio et la thèse d'exercice doit être liée à la médecine générale(40). Cela a conduit à une tendance à produire des thèses-articles de médecine générale, qui combinent les thèmes déjà étudiés dans les thèses "ancien format" avec la forme courte d'un article scientifique. Le but étant de diffuser ces thèses sous forme d'articles publiables dans des revues scientifiques, de posters, et de communications orales. Ce qui contribue à la reconnaissance scientifique des travaux des internes et des médecins généralistes.

La durée de l'internat en médecine générale en France a été prolongée de 3 ans à 4 ans à compter de 2023(41). Cette extension a été mise en place pour permettre aux internes de développer davantage leurs compétences professionnelles et de mieux se préparer à l'exercice de la médecine générale. Il a été considéré qu'avec une durée de formation plus longue, les internes pourraient bénéficier d'une plus grande expérience pratique, et d'un plus grand temps pour la formation transversale et spécialisée, pour ainsi devenir des médecins généralistes plus compétents et mieux formés. Il est noté que cette extension serait accompagnée d'un encadrement plus précis pour les internes.

La thèse d'exercice est un travail de recherche universitaire qui doit être réalisé et soutenu devant un jury pour l'obtention d'un diplôme universitaire. Historiquement, elle était nécessaire pour

obtenir le grade de docteur en médecine ou en chirurgie dans les universités françaises dès 1803, après la Révolution française. Depuis cette période, la thèse d'exercice est devenue un élément clé de la formation des médecins, permettant de valider leur capacité à mener une recherche scientifique et à contribuer au développement de la médecine.

En médecine générale, la thèse d'exercice concerne plus de 3000 doctorants annuellement. Chaque thèse est propre à chaque interne, mais y a-t-il dans tous ces travaux des éléments comparables pouvant mettre en exergue des besoins spécifiques dans la population générale ou en matière de formation initiale ?

Toutes ces thèses réalisées tous les ans sont autant d'heures de travail qu'un médecin ne passera pas auprès de ses patients. Au vu du contexte actuel de démographie médicale, et notamment au niveau de la médecine générale, nous devons essayer d'améliorer l'efficacité de la prise en soin des patients. La thèse pour un médecin généraliste va représenter plusieurs, journées, mois, voire années de travail, mais peut-on réduire cette durée en optimisant les champs à étudier en priorité et en donnant les outils aux thésards et directeurs de thèse pour échanger plus facilement ?

L'objectif principal de ce travail est de déterminer quels sujets sont prépondérants parmi les thèses de médecine générale en France en 2021 et comment ces thématiques ont évolué depuis 2015. L'objectif secondaire est de développer des méthodes d'analyse de données reproductibles et d'apprécier leurs cohérences avec les données disponibles actuellement(8,11,24,42-44) pour étudier les thématiques des thèses de médecine générale.

2 Méthode et matériel

2.1 Méthode

2.1.1 Type d'étude

Nous avons effectué une étude portant sur les thèses de médecine générale de France de 2015 à 2021. Après la recherche, l'extraction, le tri et l'uniformisation des données, les résultats ont été obtenus via des algorithmes de fouille de texte ou text mining.

2.1.2 Extraction des données

La base de données utilisée est celle de Sudoc, avec les critères de recherche avancée suivant :

- « *médecine générale* » doit être présent dans la description de la thèse
- les termes chirurgie et psychiatrie doivent être absents
- la langue sélectionnée est le français
- la période étudiée sera de 2015 à 2021
- le type de document recherché est « *thèse* ».

Cette base de données permet d'avoir accès à plusieurs milliers de fiches de thèse. Une fiche de thèse va regrouper les informations suivantes qui nous permettront de faciliter la classification future de ces dernières. La première information contenue dans ces fiches est l'identifiant pérenne de la notice, cela nous sera utile par la suite, car chaque fiche et donc chaque thèse aura un identifiant unique. Le titre, l'auteur, la date de soutenance, le type de thèse et le résumé sont aussi présents dans ces fiches.

Cette étude n'est pas visée par la loi Jardé. Toutes les données sont accessibles au public et peuvent être consultées sur Sudoc.

Les fiches peuvent être téléchargées par lot de 100, puis copiées dans un fichier texte. Ensuite, une fois que les fiches de toute une année ont été extraites, elles ont été recomptées via un algorithme comptant les identifiants des fiches de thèses pour vérifier que le nombre de fiches de thèses est le

même que celui affiché par le site Sudoc lors de la recherche. L'opération a été répétée pour chaque année étudiée de 2015 à 2021.

Le nombre de fiches de thèses extraites par année étant connu car affiché lors de la recherche dans la base de données, nous avons pu vérifier que l'algorithme triait les données correctement en comparant le nombre de fiches triées à celles attendues. La base de données comportait quelques caractères irréguliers qui interféraient avec l'algorithme de tri des données. Ces erreurs étant marginales, nous avons pu toutes les corriger et trier ainsi l'intégralité des fiches de thèses récupérées dans la base de données Sudoc.

2.1.3 Tri des données

2.1.3.1 Séparation des fiches de thèses entre elles

Ensuite une fois toutes les données récupérées, nous les avons triées, car nous voulions utiliser seulement certaines informations spécifiques. Il s'agira de la ville où la thèse est soutenue, de la date de naissance de l'auteur, de l'identifiant unique de la thèse, de l'année de soutenance de la thèse, ainsi que du résumé complet de la thèse. D'autres informations ont pu être classées mais non utilisées dans l'analyse, comme le titre de la thèse, les mots-clés utilisés pour l'indexation, le type de thèse, ou le numéro national de thèse. Pour cela nous avons écrit un programme en utilisant le langage de programmation R. Les fiches de thèses ont été séparées en utilisant encore une fois l'expression "Identifiant pérenne de la notice : ". Dès que le programme détectait strictement cette expression, le texte était alors découpé et collé dans un tableau à autant de lignes que de fiches de thèses.

2.1.3.2 Recherche et tri des données présentes dans chaque fiche

Identifiant pérenne de la notice : <https://www.sudoc.fr/195723910>

Titre : **Infection urinaire de l'adulte : prise en charge par les médecins généralistes en Guyane Française**

Auteur(s) : Briquet, Yasmina (1984-....)

Date(s) : 2016

Num. de thèse d'exercice : 2016AMIEM065

Thèse : Thèse d'exercice : Médecine. Médecine générale : Amiens : 2016

Résumé(s) : **Introduction** : Plusieurs recommandations sur le diagnostic et l'antibiothérapie des infections urinaires ont été successivement émises en France. Nous avons souhaité connaître le niveau d'adhésion des médecins généralistes guyanais à ces recommandations [...]. Les déclarations et pratiques observées restent discordantes des recommandations de 2014, notamment en ce qui concerne la réalisation des BU, des ECBU, de l'échographie et de l'antibiothérapie. Des actions de formations sous forme d'EPU sont nécessaires.

Sujets :

- Médecins généralistes -- Guyane
- Analyse d'urine
- Appareil urinaire -- Infections
- Recommandations sur le bon usage des médicaments

Figure 1 : Schéma représentant le fonctionnement du programme de tri des données

Ci-dessus un exemple d'une fiche de thèse présente dans la base de données (45). Ce schéma ne présente pas l'intégralité des informations présentes sur ces fiches. Le tri des données se fait de la façon suivante : l'algorithme va rechercher des expressions régulières, qui est en informatique une chaîne de caractères qui décrit, selon une syntaxe précise, un ensemble de chaînes de caractères possibles. Par exemple ici nous voulions récupérer l'information située derrière l'identifiant pérenne de la notice, et plus particulièrement la série de caractères derrière « <https://www.sudoc.fr/> ». Il suffit donc d'utiliser l'expression régulière « <https://www.sudoc.fr/>`[[^]{9}` » qui nous permet d'extraire cette expression plus neuf caractères derrière celle-ci. Nous avons ainsi récupéré tous les identifiants uniques des fiches de thèses.

En utilisant une méthode similaire à cette dernière, nous pouvons donc extraire les informations qui nous intéressent, en rouge sur la Figure 1, en utilisant comme repère les expressions vertes ainsi que quelques expressions régulières. Puis nous avons pu supprimer les informations apparaissant de manière systématique liées directement à la rédaction de la fiche de thèse et n'ayant pas d'intérêts à être analysées par la suite étant donné que ces informations sont présentes sur toutes les fiches.

Ces opérations nous permettent de créer une base de données tabulaire avec une thèse par ligne et les différentes variables d'intérêts en colonne.

2.1.3.3 Harmonisation de certains lieux d'origine des fiches de thèse

Tableau 1: Liste des noms de villes remplacés

Nom de l'université dans la base de données	Remplacé par
Aix Marseille	Aix-Marseille
Université de Lorraine	Nancy
Saint-etienne	Saint-Etienne
Saint-Étienne	Saint-Etienne
UPEC	Paris
Université de Paris	Paris
Université Grenoble Alpes	Grenoble
Sorbonne Université	Paris
Versailles-St Quentin en Yvelines	Paris

Une homogénéisation des dénominations des UFR a été effectuée selon le tableau suivant. Cette homogénéisation permet de tenir compte des changements de noms de certaines UFR sur la période d'étude. Les changements suivants ont été effectués quant à la dénomination des UFR de médecine. Ils permettent de faciliter la lisibilité des résultats, pour pallier les changements de noms de certaines universités durant la période d'étude et pour que l'utilisation de caractères spéciaux n'interfère pas dans l'exploitation des données.

2.1.3.4 Exclusion des fiches de thèses ne contenant pas de résumés

La recherche des thématiques des thèses repose essentiellement sur l'analyse des résumés de celles-ci. Or le résumé n'est pas toujours présent dans la base de données, les fiches de thèses présentant moins de 100 caractères ont été exclues de la population étudiée. Ce seuil a été choisi arbitrairement pour définir si un résumé est présent ou non. Il a été décidé que si moins de 100 caractères étaient retrouvés dans la section résumé, celle-ci ne pouvait pas être pertinente vis-à-vis du travail de thèse.

2.1.4 Exploitation des données globales

2.1.4.1 Estimation de l'âge des auteurs et de l'âge moyen de passage de la thèse

Une estimation de l'âge des auteurs a pu être calculée en soustrayant l'année de naissance de chaque auteur à l'année de soutenance de la thèse correspondante. Une moyenne par année a ensuite été effectuée permettant d'obtenir une estimation de l'âge moyen de passage de la thèse en fonction des années étudiées.

2.1.4.2 Répartition géographique de l'échantillon

La répartition géographique a été déterminée via le dénombrement des origines des fiches de thèses extraites. Puis chaque origine a ensuite été représentée sur une carte de France, la taille est proportionnelle à l'effectif.

2.1.4.3 Effectif et proportion des thèses analysées par années

L'effectif et la proportion des fiches de thèses par année ont pu être calculés en dénombrant le nombre de thèses par année ainsi que le nombre total de fiches de thèses récupérées.

2.1.5 Préparation du corpus en vue de son analyse

2.1.5.1 *Mise en forme de la base de données*

Pour la suite des analyses, nous ne nous intéresserons plus qu'aux données suivantes : le résumé extrait des fiches de thèses, son identifiant unique, ainsi que l'année de soutenance. Un preprocessing a été effectué sur les résumés des thèses. Tous les caractères du corpus sont transformés en minuscules, la ponctuation ainsi que les espaces vides superflus sont supprimés.

2.1.5.2 *Lemmatisation*

L'algorithme d'analyse n'est pas capable de différencier les formes grammaticales des mots, même s'ils représentent la même idée. Par exemple le terme « médecin » sera perçu comme différent du terme « médecins ». Pour pallier cela, nous utilisons la lemmatisation, qui consiste à remplacer un mot par sa forme canonique. Il existe des bibliothèques le faisant automatiquement, mais comme cette action dépend parfois du contexte de la phrase cela peut conduire au remplacement d'un terme qui change son sens. Par exemple un des problèmes présents via cette méthode est que le terme « CHU » (centre hospitalo-universitaire) était systématiquement remplacé par le verbe « choir ». Nous avons donc choisi d'effectuer une lemmatisation manuelle pour éviter ces problèmes, la liste des termes ainsi remplacés se trouve en Annexe 1, p143.

2.1.5.3 *Racinisation*

La racinisation permet de ne garder que la racine d'un mot, et ainsi regrouper plusieurs mots d'une même famille. Cependant le fait de raciner les mots peut rendre leur interprétation difficile, c'est pourquoi en première partie d'analyse nous n'avons pas utilisé cet outil, car nous nous intéressions à des mots uniques. Nous avons cependant utilisé cet outil en dernière partie d'analyse où nous avons regroupé tous les termes pouvant appartenir au même thème.

2.1.5.4 *Suppression des mots vides ou stopwords*

En recherche d'information, un mot vide est un mot qui est tellement commun qu'il est inutile de l'indexer ou de l'utiliser dans une recherche (5). Nous avons donc supprimé ces mots en nous aidant

d'un dictionnaire de mots vides pour la langue française et anglaise. Ce sont bien souvent des mots très communs comme par exemple des articles, qui n'apportent pas d'information en tant que telle. La liste complète de ces mots qui ont pu être supprimés du texte original est disponible en Annexe 2, p145.

2.1.6 Occurrences totales des mots

Après ces différentes transformations des textes du corpus, nous pouvons effectuer plus aisément les analyses nécessaires. La première analyse effectuée est un décompte de tous les mots avec un classement des mots les plus fréquents. Comme l'année de soutenance de la thèse a été conservée dans les données, il est aussi possible d'afficher le dénombrement de ces mots les plus fréquents par années.

2.1.7 Décompte des paires de mots au sein d'un même document

Cet outil nous permettra de décompter le nombre de fois où un couple de mots apparaîtra au sein d'un même résumé, et ce sur l'ensemble du corpus. Il permettra donc de retrouver les mots fréquemment utilisés ensemble ainsi que leur nombre et leur fréquence.

2.1.8 Corrélation entre les différents mots

Le décompte des paires de mots peut ne pas être significatif, car cela peut être des mots très communs individuellement. Il est possible de calculer la corrélation des termes entre eux, c'est-à-dire la fréquence à laquelle ces mots apparaissent ensemble par rapport à la fréquence où ils apparaissent séparément. Cela permettra de savoir si des termes sont fortement liés et apparaissent souvent ensemble.

Tableau 2 : Calcul de la corrélation entre les termes

	Présence du mot Y	Absence du mot Y	Total
Présence du mot X	n_{11}	n_{10}	$n_{1.}$
Absence du mot X	n_{01}	n_{00}	$n_{0.}$
Total	$n_{.1}$	$n_{.0}$	N

Équation 1 : Corrélation entre les termes

$$\varphi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1.}n_{0.}n_{.1}n_{.0}}}$$

Avec cet outil il est possible d'établir des représentations graphiques des corrélations entre les termes. Seuls les termes apparaissant plus de vingt fois en 2021 ont été pris en compte. Les résultats ont ensuite été filtrés en supprimant des mots n'ayant pas de corrélations avec d'autres ou étant vides de sens afin de faciliter la lisibilité. Ces mots sont consultables en annexe 2.

2.1.9 La fréquence de terme et la fréquence de document inverse

La fréquence de terme et la fréquence de document inverse ou Terme Frequency et Inverse Document Frequency abrégé en TF-IDF, permet de pondérer les résultats obtenus dans la fouille de texte. Il permettra de dégager des termes qui sont très représentés seulement dans quelques documents, des termes présents dans tous les documents. Plus le terme sera présent dans un document du corpus plus sa valeur de fréquence (TF) va augmenter, mais comme elle est pondérée par la valeur de la fréquence de document inverse, plus ce terme est présent dans beaucoup de documents, plus sa valeur de TF-IDF sera faible. Cela permet en pratique de trouver des termes représentatifs des sujets abordés tout en occultant les termes utilisés très fréquemment pouvant parasiter l'analyse (6).

Équation 2 : Calcul du TF

$$TF = \frac{n_x}{\sum_i^j n_{i,j}}$$

Ci-dessus le calcul pour la fréquence de terme, il s'agit de compter le nombre d'occurrences où le terme X apparaît dans un document divisé par la somme de tous les termes présents dans ce document.

Équation 3 : Calcul de l'IDF

$$IDF = \log \frac{|D|}{|N|}$$

Ci-dessus le calcul de l'IDF qui est le logarithme du nombre total de documents contenus dans le corpus étudié sur le nombre de documents où le terme X apparaît. Le calcul du TF-IDF s'obtient ensuite par multiplication des deux termes obtenus, notons que l'IDF ne change pas pour un terme étudié dans le corpus alors que le TF va être propre à chaque document.

Donc d'après cette formule, on peut voir que si un terme est présent dans tous les documents, son IDF sera égal à $\log \frac{1}{1} = 0$, donc son TF-IDF le sera aussi.

Il est ensuite possible de calculer les corrélations entre ces différents termes comme fait précédemment.

2.1.10 Analyse en bigramme

Le bigramme est une paire de mots qui se suivent, il est ainsi possible d'analyser l'ensemble du corpus non pas sur chaque terme, mais sur des couples de termes. Par exemple dans la phrase « Il fait beau aujourd'hui » le texte sera découpé en : « Il fait », « fait beau », « beau aujourd'hui » et analysé en tant que tel. Cela permet de récupérer des informations supplémentaires quant à la syntaxe qui n'est pas possible en analysant mot par mot. Ici nous nous sommes limités à deux termes pour l'analyse, mais en fonction des besoins il est possible de construire des groupes de plusieurs mots, ou d'un nombre de caractères précis. À noter que la puissance de calcul nécessaire augmente en fonction du nombre de mots ou de caractères que l'on souhaite analyser. Sur notre exemple nous pouvons voir qu'une phrase de quatre mots est transformée en six mots au total. Il est bien sûr possible de poursuivre l'analyse de ces bigrammes en utilisant les outils précédents, par exemple l'analyse de corrélation des bigrammes des résumés de thèses que nous avons réalisées, avec la liste des mots vides supprimés de l'analyse disponible en annexe 3

2.1.11 Approche par apprentissage automatique.

Jusqu'à présent, nous avons cherché à dégager des termes significatifs par divers outils afin de pouvoir modéliser d'éventuels sujets. Mais il est possible d'utiliser des algorithmes qui classeront

automatiquement les termes dans les sujets correspondants. Les deux principales méthodes utilisées seront l'échantillonnage de Gibbs et l'allocation de Dirichlet latente.

2.1.11.1 Échantillonnage de Gibbs

Le principal problème est que nous ne connaissons pas, à priori, le nombre de sujets possibles pour notre corpus, il est bien sûr possible de l'estimer empiriquement, mais il est également possible de calculer le nombre de sujet optimal pour la classification.

L'échantillonnage de Gibbs va partir de l'hypothèse qu'il n'existe qu'un sujet, puis essayer de classer tous les termes dans cet ensemble et en déterminer la probabilité. Puis de façon incrémentielle l'algorithme va partir de l'hypothèse qu'il existe deux sujets, puis trois, etc. De cette façon on pourra déterminer pour quel nombre de sujets cette probabilité est la plus importante et sélectionner cette valeur pour la suite. Nous allons utiliser quatre fois cet échantillonnage selon quatre méthodes de calcul différentes (47–50) et déterminer le nombre de sujets d'après ces quatre approches simultanément.

2.1.11.2 Allocation de Dirichlet latente

C'est un modèle probabiliste génératif (51) qui va nous servir à classer chaque document afin qu'il tende vers un seul sujet par document et que chaque mot soit relatif à un sujet. L'algorithme va ainsi regrouper tous les termes qui auront une probabilité importante d'apparaître ensemble, en déduire le ou les sujets de chaque document. En revanche comme dit précédemment le modèle ignore le nombre de sujets potentiels, il faut donc soit le choisir empiriquement, soit s'aider de l'échantillonnage de Gibbs par exemple. Une fois le nombre de sujets estimé, l'algorithme va les classer en se basant sur ce nombre, en différents sujets. Le résultat final sera donc plusieurs groupes de termes qui sont en relation avec un sujet. La teneur du sujet n'est en revanche pas identifiable de manière automatique, il faudra donc le déduire d'après ces groupes de mots.

2.1.12 Évolution des sujets de thèses de 2015 à 2021

2.1.12.1 *Analyse par fréquence de terme dans le corpus*

Tout d'abord nous allons écarter de l'analyse les thèmes retrouvés par LDA faisant référence au type ou à la méthodologie d'étude. Nous allons choisir un ou plusieurs termes représentatifs d'un thème ou de plusieurs thèmes, puis nous allons détecter ce ou ces termes dans tout le corpus. Cela nous permettra de retrouver combien de résumés de thèses par année contiennent ce terme. Un modèle de régression linéaire sera appliqué, afin de pouvoir mieux apprécier son évolutivité dans le temps. L'évolution dans le temps de certains thèmes ne pourra pas être estimé par cette méthode, car ces derniers auront un champ lexical trop général pour avoir une représentation de leur évolution cohérente. De même il n'est pas possible de juger si la thématique est principale ou secondaire dans le résumé. Ensuite, nous joindrons l'analyse des dix mots ayant la corrélation la plus importante avec le terme choisi, afin de pouvoir confirmer ou infirmer que ce terme soit représentatif du sujet dont nous voulons explorer l'évolution. À noter que pour faciliter l'analyse, nous avons raciné le corpus comme pour l'analyse en LDA, nous avons ensuite lemmatisé les formes racines pour améliorer la lisibilité et la compréhension des résultats.

2.1.12.2 *Analyse et comparaison de l'évolution par LDA*

Nous allons refaire un classement des sujets par LDA en prenant cette fois-ci l'intégralité du corpus. Le désavantage est que nous ne retrouverons pas de façon invariable toutes les mêmes thématiques trouvées par LDA sur l'année 2021 seule, étant donné que certaines sont apparues ou ont diminué de 2015 à 2021.

En revanche cela va permettre de pouvoir apprécier l'évolution de thématiques principales ainsi que celles ayant un champ lexical plus général.

2.2 Matériel

Le matériel utilisé est un ordinateur muni d'une connexion internet, d'un processeur Intel i7-7700K et d'une mémoire vive de 32 Go. Le système d'exploitation utilisé est Windows 10 professionnel et le logiciel pour la fouille de donnée utilisé est R version 4.1.0.

Le matériel utilisé pour le tri et le traitement des données peut avoir une incidence sur les résultats obtenus, allant même jusqu'à l'impossibilité de réaliser certaines opérations si la puissance de calcul ou la mémoire disponible n'est pas suffisante.

3 Résultats

3.1 Données globales de l'échantillon

3.1.1 Taille de l'échantillon et période d'étude retenue

Concernant notre échantillon, 11 729 fiches de thèses ont été extraites de la base de données Sudoc. Après tri des données, 11 026 fiches de thèses ont été conservées pour l'analyse des données. La période d'étude s'étend de 2015 à 2021, elle englobe l'année de la réforme du troisième cycle des études médicales en France (52).

3.1.2 Répartition géographique de l'échantillon



Figure 2 : Répartition géographique du nombre de fiches de thèses analysées

Tableau 2 : Répartition géographique des fiches de thèse récupérées

Ville	Effectif	%
Paris	1668	14,22
Lille	1506	12,84
Bordeaux	1420	12,11
Aix-Marseille	1025	8,74
Toulouse	782	6,67
Nantes	722	6,16
Poitiers	643	5,48
Saint-Etienne	532	4,54
Amiens	527	4,49
Grenoble	457	3,9
Strasbourg	411	3,5
Rouen	387	3,3
Reims	323	2,75
Limoges	245	2,09
La Réunion	156	1,33
Nancy	86	0,73
Angers	65	0,55
Lyon	52	0,44
Indéterminée	10	0,09
Dijon	6	0,05
Nice	2	0,02
Tours	1	0,01

Nous constatons que, sur la Figure 2 et le Tableau 2 ci-dessus, les principales villes représentées sont Paris à 14,22 %, Lille à 12,84%, Bordeaux à 12,11% et Marseille à 8,74%. D'autre part, sur la totalité de l'échantillon, il n'y a que 10 fiches de thèse pour lesquelles l'origine n'était pas spécifiée.

3.1.3 Âge des auteurs lors de la soutenance de la thèse

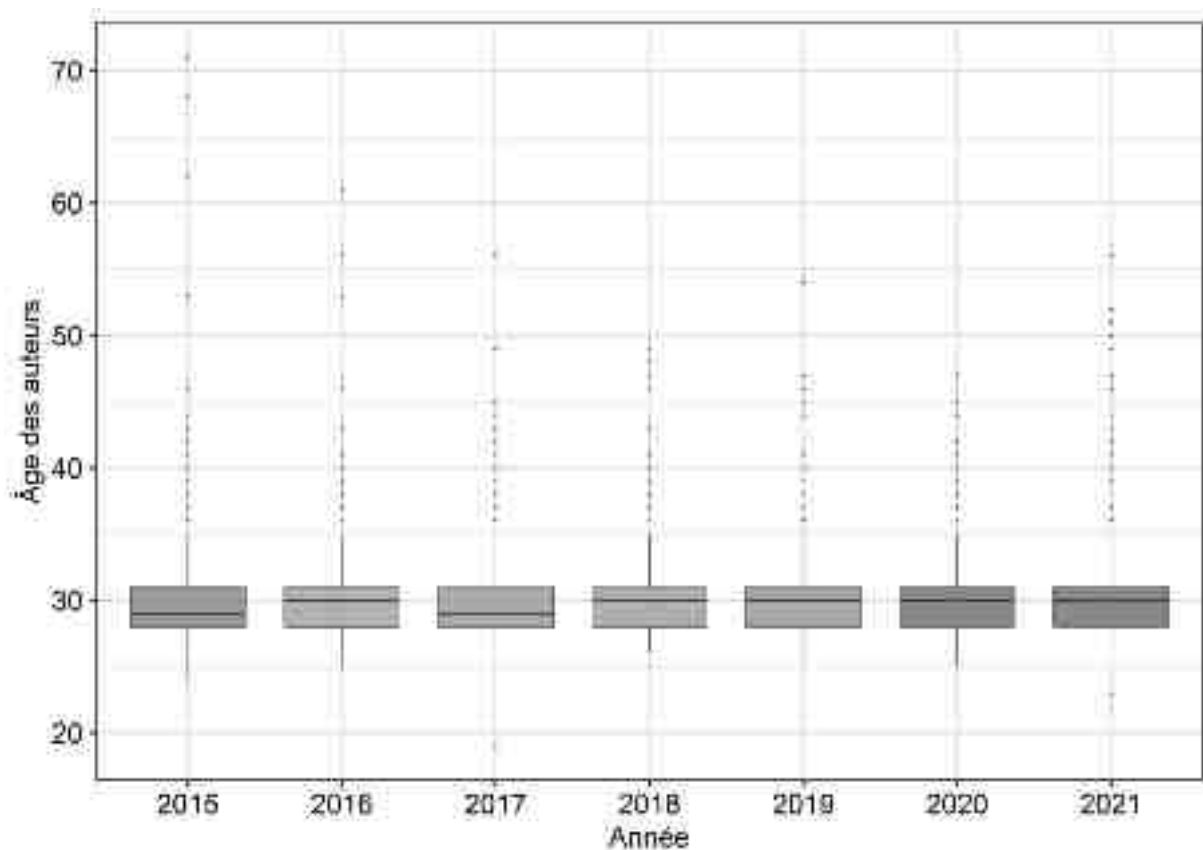


Figure 3 : Âge des auteurs lors de la soutenance de thèse

Sur la Figure 3, nous observons la répartition des âges de soutenance de l'année en fonction de l'année. Nous constatons que la plupart des valeurs se situent entre 28 ans et 31 ans, et semblent régulières dans le temps. Les valeurs extrêmes sont pour la plupart situées au-dessus de 35 ans et jusqu'à plus de 70 ans (cette dernière valeur a été vérifiée et il ne s'agit pas d'une erreur). Les valeurs extrêmes en dessous de 25 ans sont beaucoup plus rares, et celle de 2017 qui affiche un âge de moins de 20 ans est en réalité une erreur à l'intérieur de la base de données Sudoc, que nous n'avons pas corrigée pour illustrer un des types de biais.

3.1.3.1 Âge des auteurs lors de la soutenance de la thèse

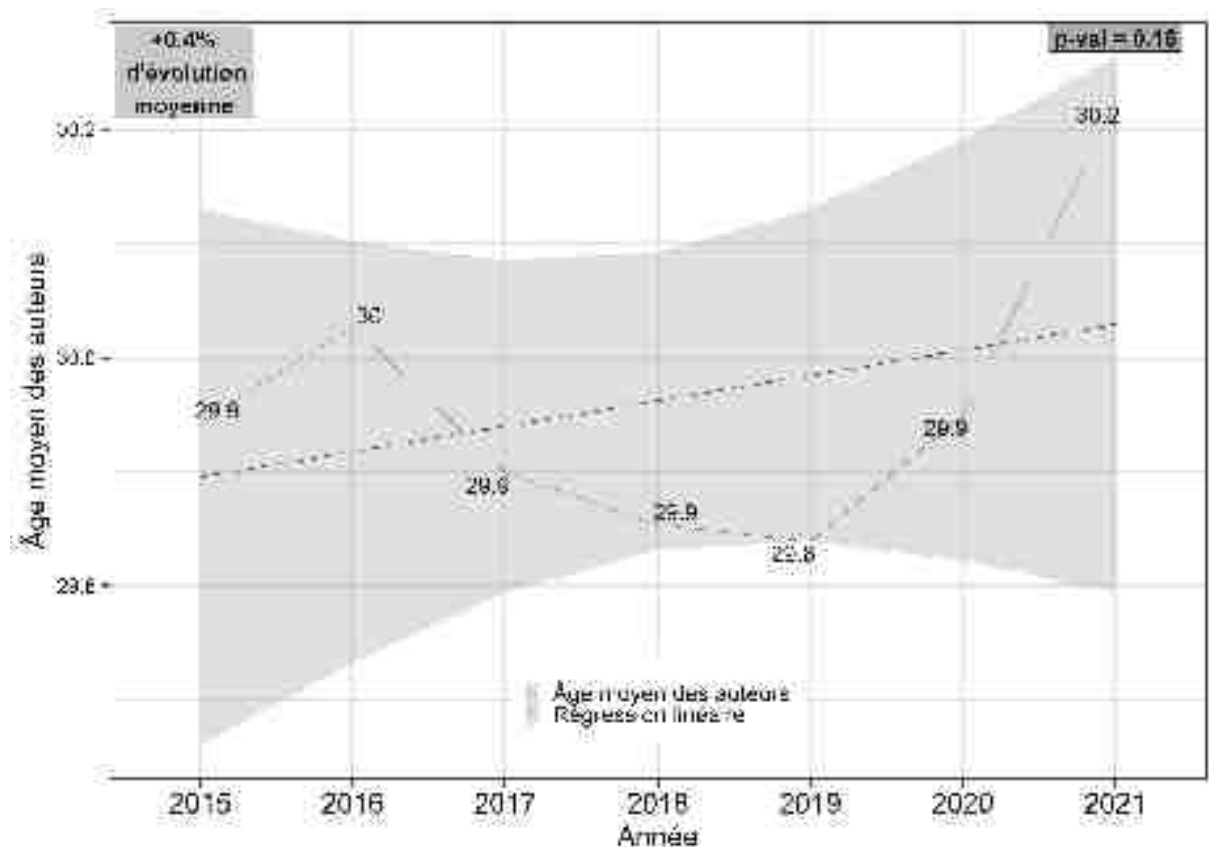


Figure 4 : Âge moyen des auteurs lors de la soutenance de thèse

Sur la Figure 4, nous avons calculé l'âge moyen des auteurs en fonction des années. L'âge moyen semble être en augmentation sur l'année 2021 où il atteint un maximum de 30,2 ans, soit +0,3 ans de 2020 à 2021. L'écart entre deux années consécutives n'était auparavant jamais plus élevé que 0,1 ans. La modélisation linéaire renvoie une évolution de l'âge moyen sur la période 2015 à 2021 de +0,4%. La différence d'âge entre les années n'est pas particulièrement importante (- de 1 an)

3.1.4 Effectif des fiches de thèses par années

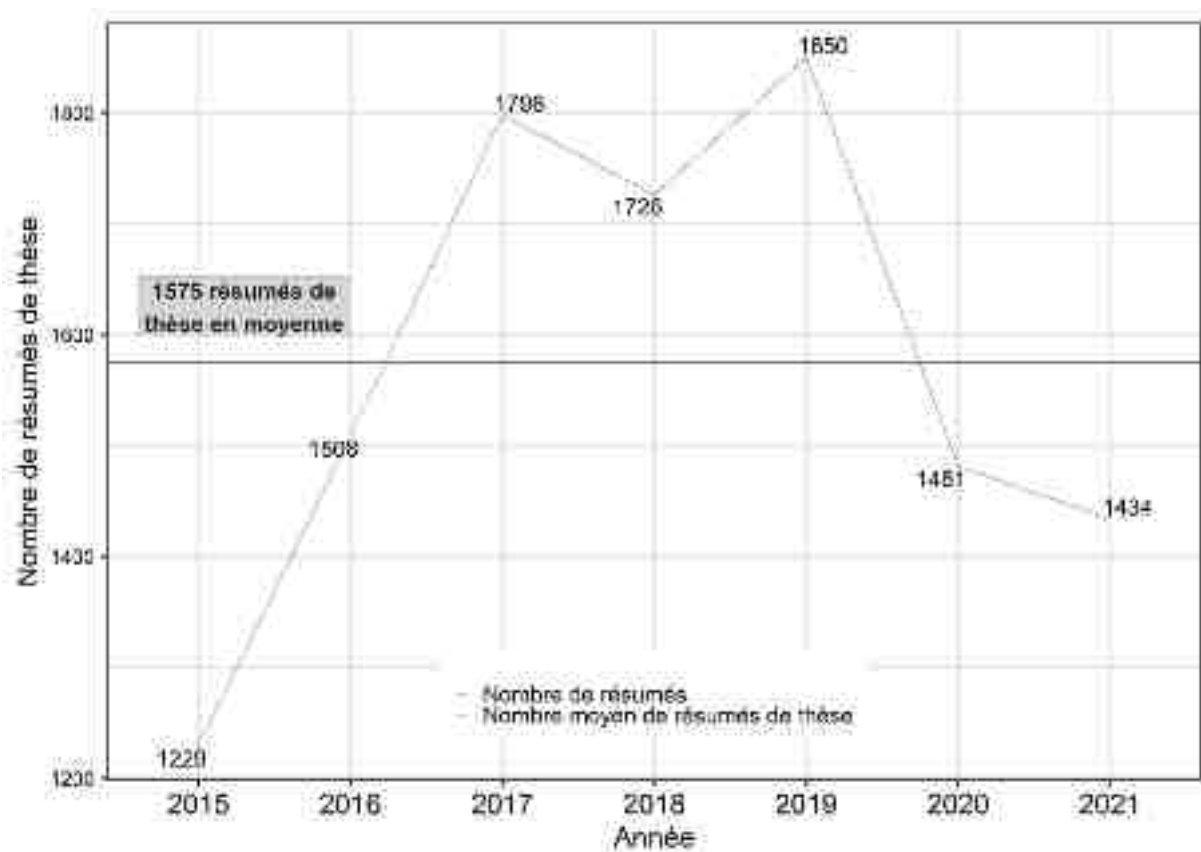


Figure 5 : Représentation graphique du nombre de fiches de thèses récupérées par années

Figure 5, nous observons le nombre de résumés de thèses récupérés par années. En prenant en compte ces statistiques sur les places attribuées aux ECN en médecine générale de l'année 2012 à 2018 (53), nous pouvons établir que 3497 places de médecine générale ont été pourvues chaque année en moyenne. Si 100 % des étudiants ayant choisi l'internat de médecin générale le finissent, nous pouvons estimer notre échantillon comme représentant environ 45 % de la population totale en moyenne.

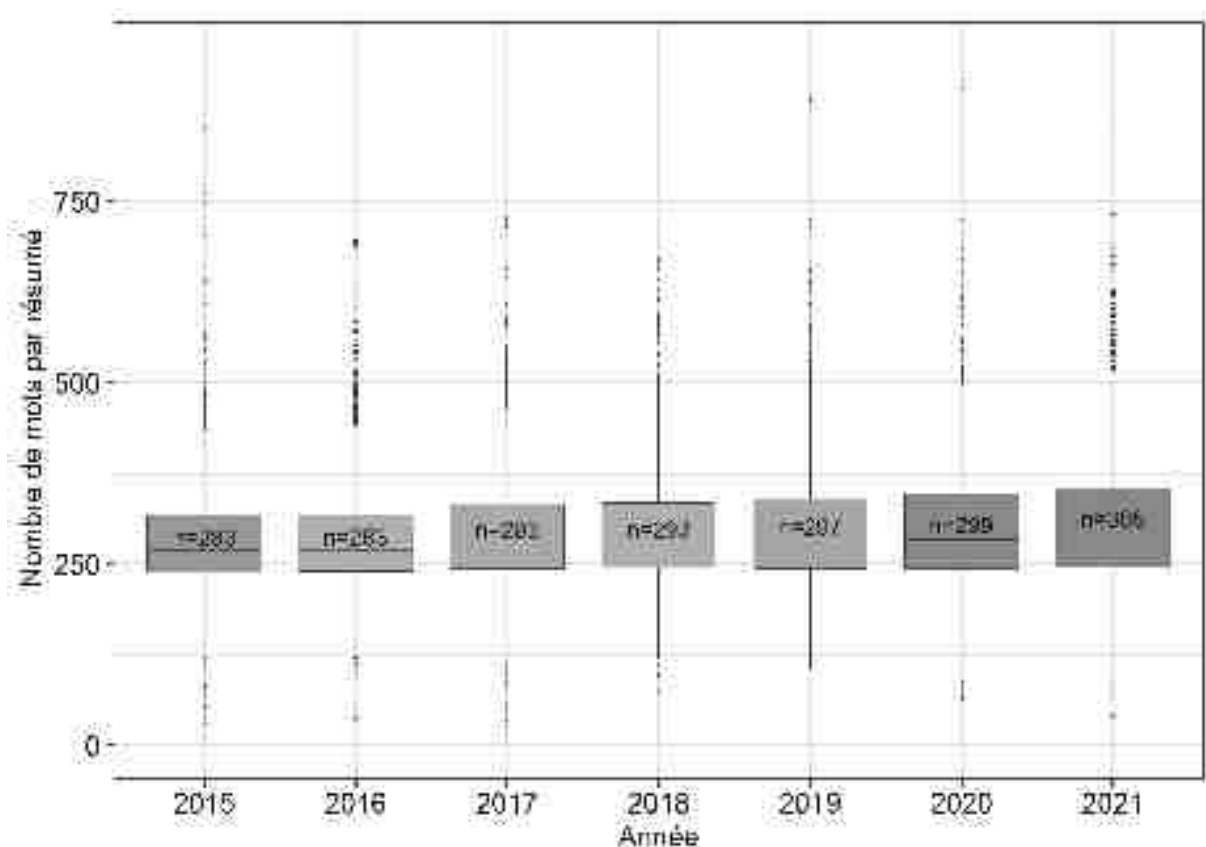


Figure 6 : Nombre de termes par résumé

Sur la Figure 6, nous avons effectué un décompte de tous les mots avant transformation du corpus. Nous observons des valeurs moyennes, en augmentation régulière de 283 mots en moyenne en 2015 à 308 mots en 2021. Par ailleurs, en 2020 nous identifions quelques valeurs extrêmes, dont l'une, à plus de 800 mots. Après vérifications dans le corpus il ne s'agit pas d'erreur de tri, mais de résumés particulièrement longs présents dans la base de données.

3.2 Analyse par fouille de texte

3.2.1 Décompte de tous les termes

3.2.1.1 Décompte de tous les termes toutes années confondues



Figure 7 : Représentation des mots les plus fréquents par nuage de mots

Figure 7, nous observons que sur les vingt termes les plus fréquents sur la période étudiée, les résultats sont trop peu précis pour mettre en évidence un potentiel sujet.

3.2.1.2 Décompte de tous les termes par année

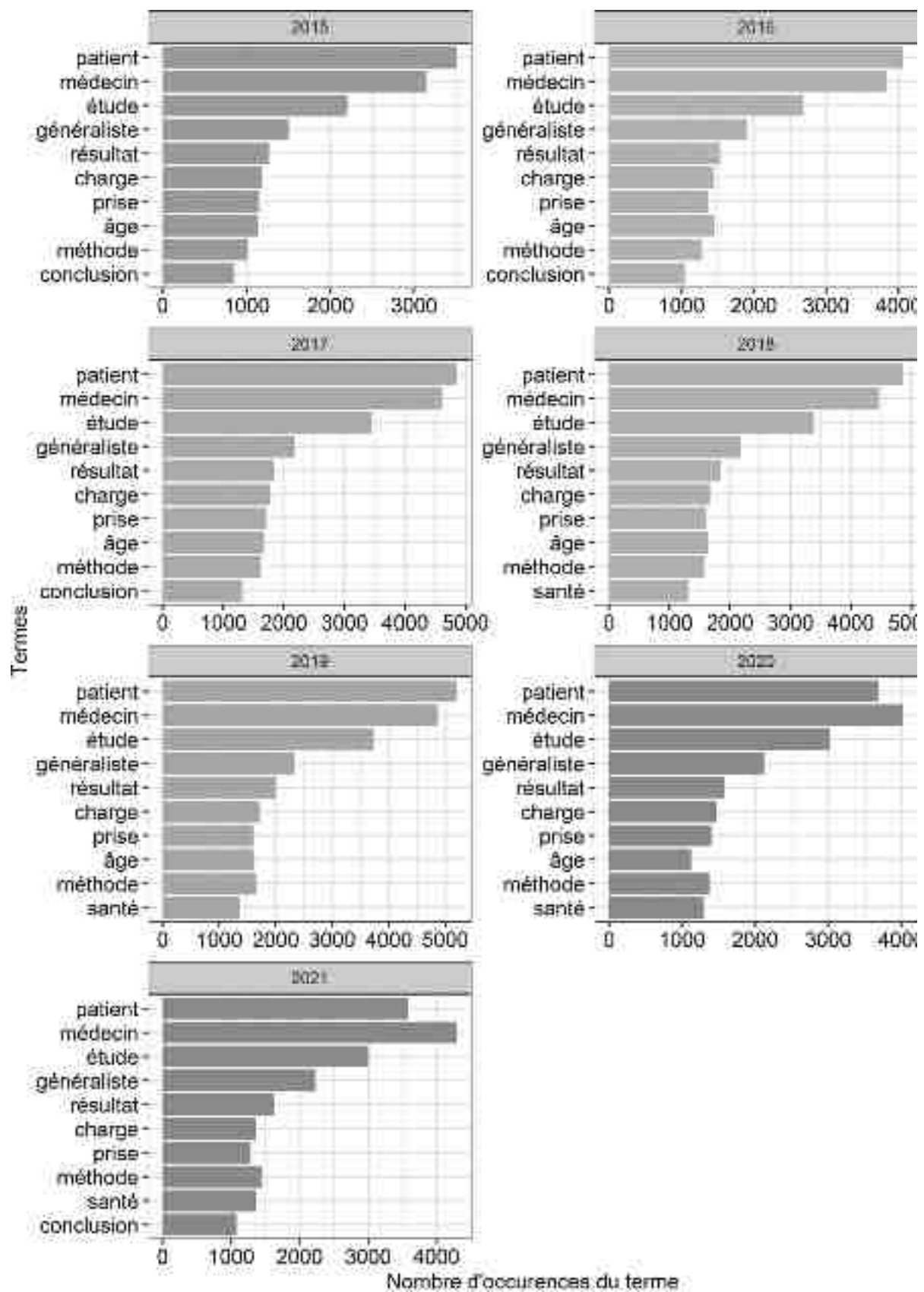


Figure 8 : Nombre d'occurrences des 10 termes les plus fréquents par année

Sur la Figure 8, nous observons que les termes les plus fréquents sont similaires. Nous avons néanmoins constaté sur les années 2020 et 2021 que la proportion du terme « médecin » est plus importante que celle de « patient ». Nous nous sommes donc interrogés sur ce résultat.

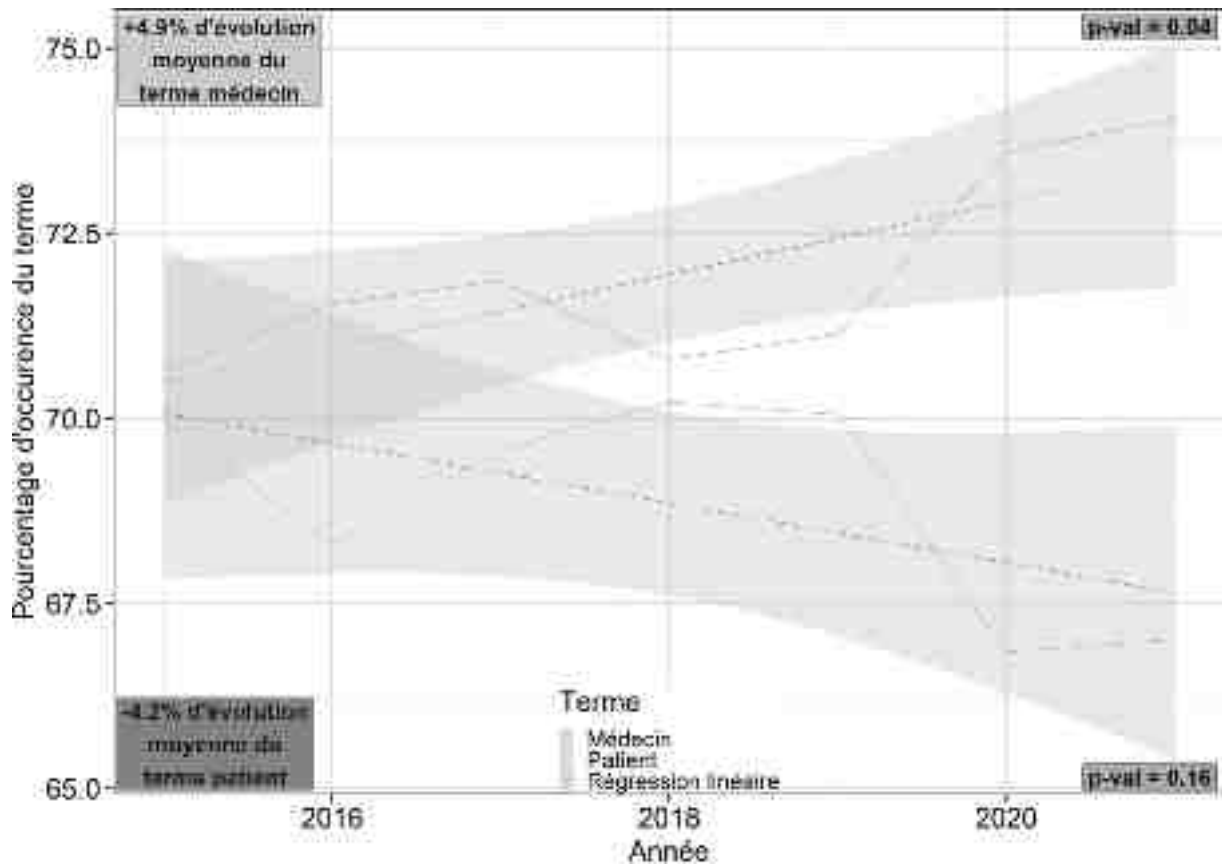


Figure 9 : Présence des termes "patient" et "médecin" dans chaque résumé de thèse

Figure 9, nous observons la présence dans les résumés de thèses des termes « médecin » et « patient » en fonction des années. Nous remarquons que la présence du terme « médecin » augmente de 4,9% de manière significative ($p=0,04$) alors que le terme « patient » lui semble être en diminution de 4,2 %, mais de façon non significative ($p=0,16$).

3.2.1.3 Décompte des termes sur l'année 2021

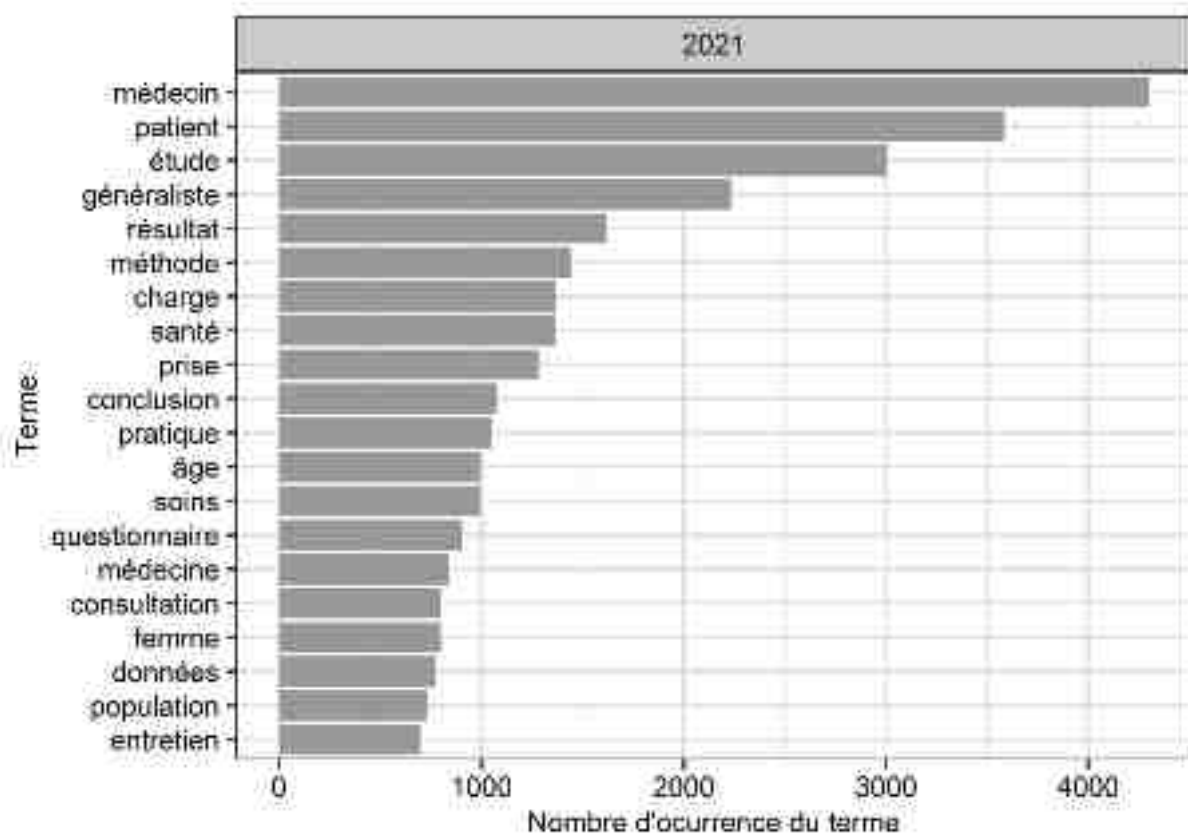


Figure 10 : Les 20 termes les plus cités dans les résumés de thèses en 2021

Afin de pouvoir déterminer quelle est l'évolution de certains sujets sur la période 2015 à 2021, nous allons devoir d'abord identifier des thèmes prépondérants sur une année en particulier. Ici nous avons choisi l'année de 2021, donc la plus récente, car les résultats obtenus seront plus utiles pour les futurs thésards et professeurs. Mais nous aurions tout aussi pu partir de l'année 2015 et en identifier l'évolution de ces sujets sur cette période, dans une approche plus centrée sur l'histoire de la médecine.

Pour commencer notre analyse des résumés de thèses de cette année nous avons décidé d'élargir le nombre de termes les plus fréquents à 20. Sur la Figure 10, le terme « femme » semble être spécifique à l'ensemble des sujets. Malgré tout, en utilisant seulement le décompte des mots contenus

dans les résumés de thèses, il nous paraît évident que les résultats obtenus sont beaucoup trop généraux pour identifier un début de réponse à notre travail de recherche.

3.2.2 Fréquence d'association des mots entre eux en 2021

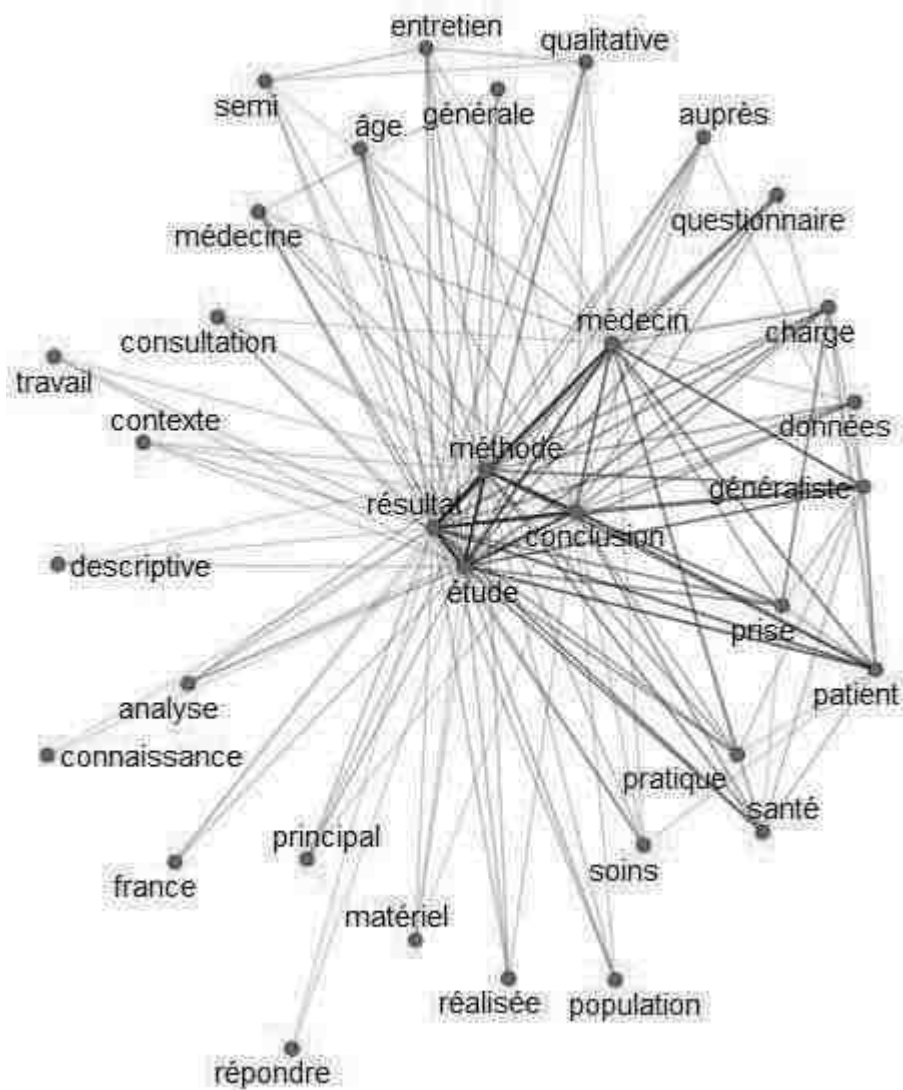


Figure 11 : Fréquence d'association des termes entre eux

Figure 11, nous observons le résultat de l'analyse de l'association des mots entre eux par paires, sur l'année 2021 et en ne prenant en compte que les termes apparaissant plus de 300 fois.

Les mots les plus proches sur le graphique sont ceux qui apparaissent le plus souvent ensemble au sein des résumés de thèses, l'épaisseur du trait apporte aussi une information sur l'importance de la

cooccurrence des mots. Nous observons qu'encore une fois cette méthode n'est pas suffisamment précise pour permettre d'identifier les sujets abordés dans ces thèses. Elle met néanmoins en évidence que la structure « méthode, résultat conclusion » est très représentée et que la thématique globale est cohérente avec les thèses de médecine générale.

3.2.3 Corrélation des termes entre eux

3.2.3.1 Corrélation des termes entre eux sans tri des termes

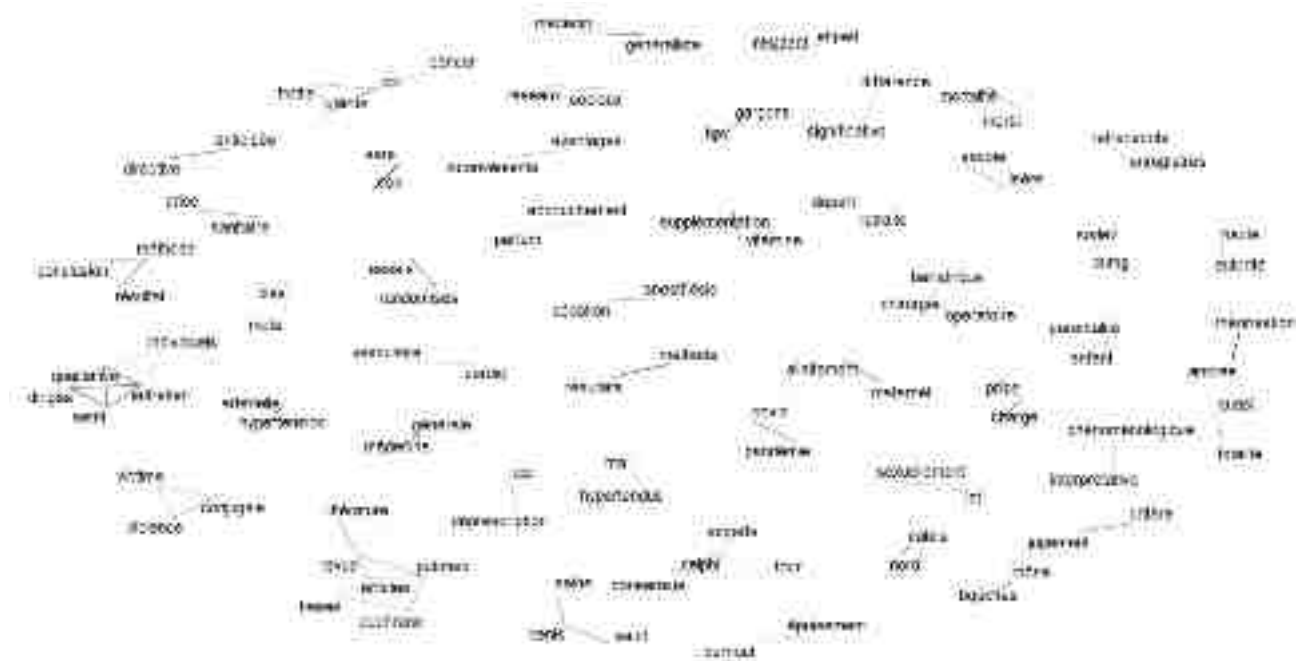


Figure 12 : Corrélation de plus de 50% entre les termes apparaissant au moins 20 fois dans les thèses de 2021 sans tri

Sur la Figure 12, nous observons les termes qui apparaissent au moins 20 fois dans le corpus et qui ont une corrélation de plus de 50 % entre eux. Ils sont représentés ici par une ligne entre deux termes. Cette méthode permet de faire davantage ressortir des idées précises. Par exemple observer une association « covid » et « pandémie », entre « épuisement » et « burnout » et entre « victime », « violence » et « conjugale ». Nous pouvons par ailleurs observer que certains termes seront a priori peu significatifs pour déterminer le sujet de ces résumés. Par exemple les lieux, la méthodologie ainsi que certaines expressions utilisées très fréquemment dans ces études ne seront pas déterminants pour rechercher des thèmes récurrents dans le corpus.

- la maternité,
- la pédiatrie et notamment les rapports aux écrans,
- la prise en charge de l'obésité.

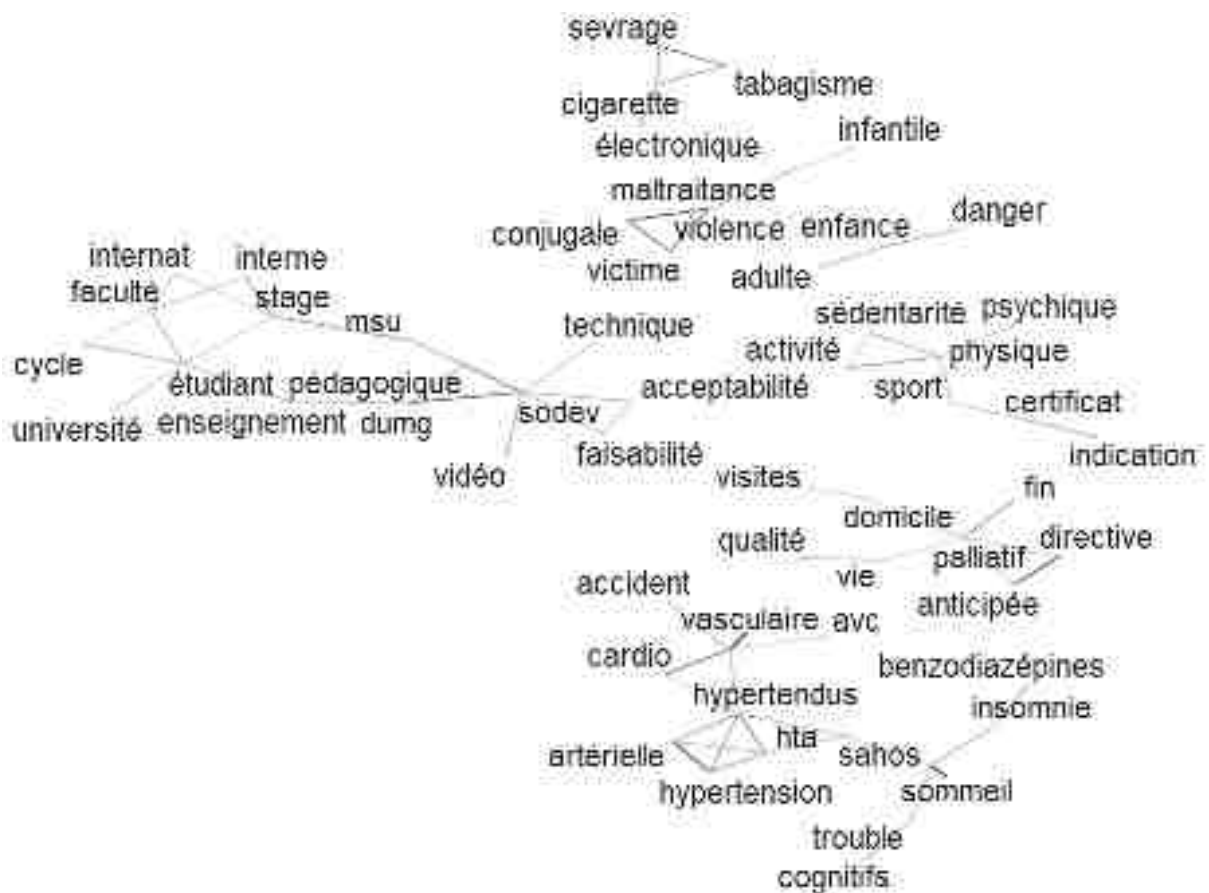


Figure 15 : Agrandissement du quadrant supérieur gauche de la Figure 13

Après avoir observé une autre zone de la Figure 13, (Figure 15) nous observons un groupe relatif :

- à la pédagogie,
- aux maladies cardiovasculaires et au sommeil,

- à la médecine à domicile et aux soins palliatifs,
- à l'activité physique,
- au sevrage tabagique avec la cigarette électronique,
- aux violences avec l'apparition d'une corrélation avec la maltraitance infantile.

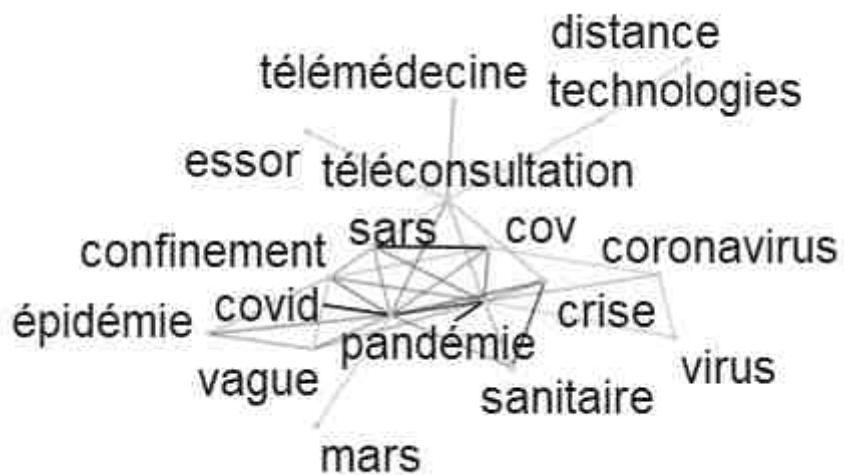


Figure 16 : Agrandissement du quadrant inférieur gauche de la Figure 13

La Figure 16 concentre des termes relatifs à la pandémie de COVID 19 ainsi qu'à la télémedecine.

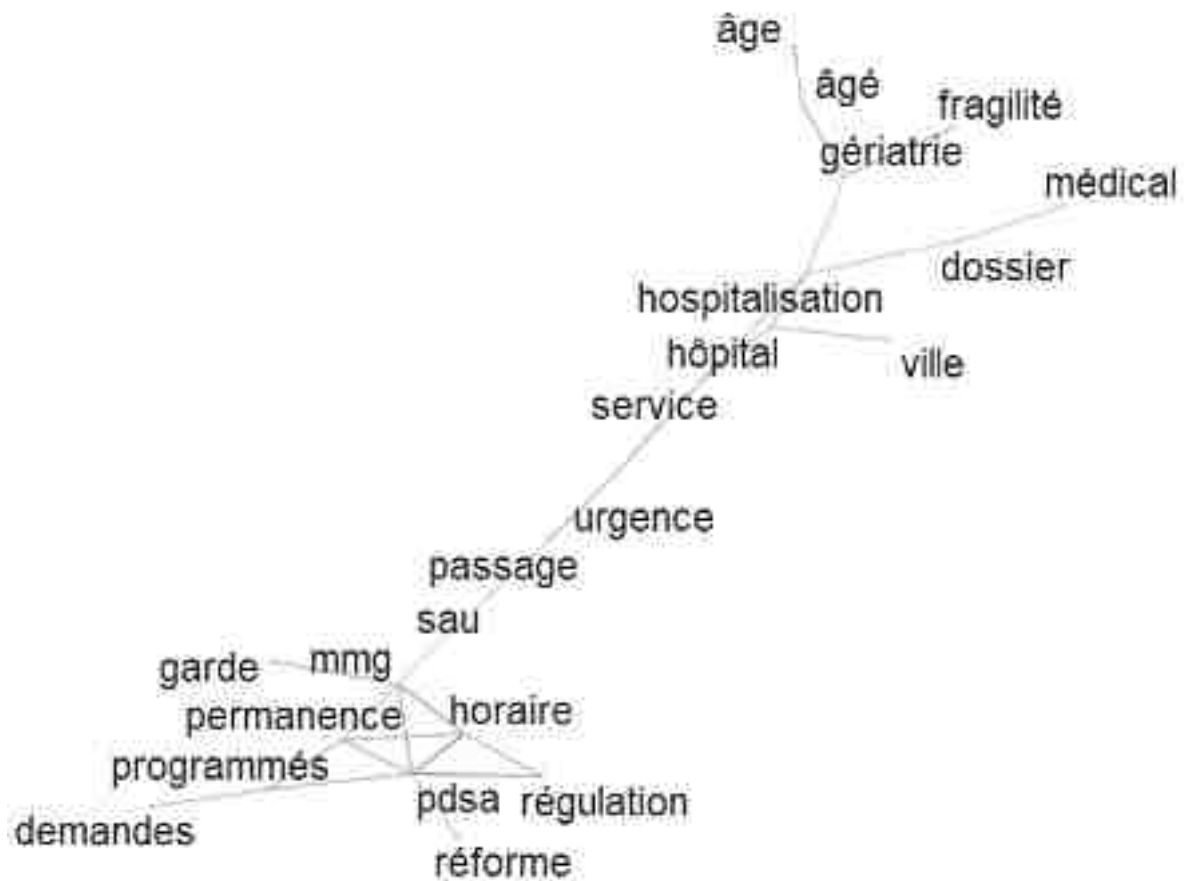


Figure 17 : Agrandissement du quadrant inférieur de la Figure 13

Sur la Figure 17 les termes seront en relation avec la permanence de soins, les urgences, les hospitalisations et la gériatrie.

3.2.4 Fréquence de terme et fréquence de document inverse



Figure 18 : Les 20 termes ayant le TF-IDF le plus grand en 2021

Le calcul du TF-IDF permet de retrouver les termes ci-dessus en Figure 18. À noter que ces derniers sont similaires à ceux obtenus en Figure 13. Seuls quelques termes « IVG », « chute » ou « IPP » n'étaient pas clairement ressortis via le calcul de corrélations des termes entre eux. Ces constats démontrent l'importance de la pondération des résultats par l'IDF. En effet les termes qui apparaissent dans tous les résumés sont mathématiquement écartés des valeurs de TF-IDF les plus hautes.

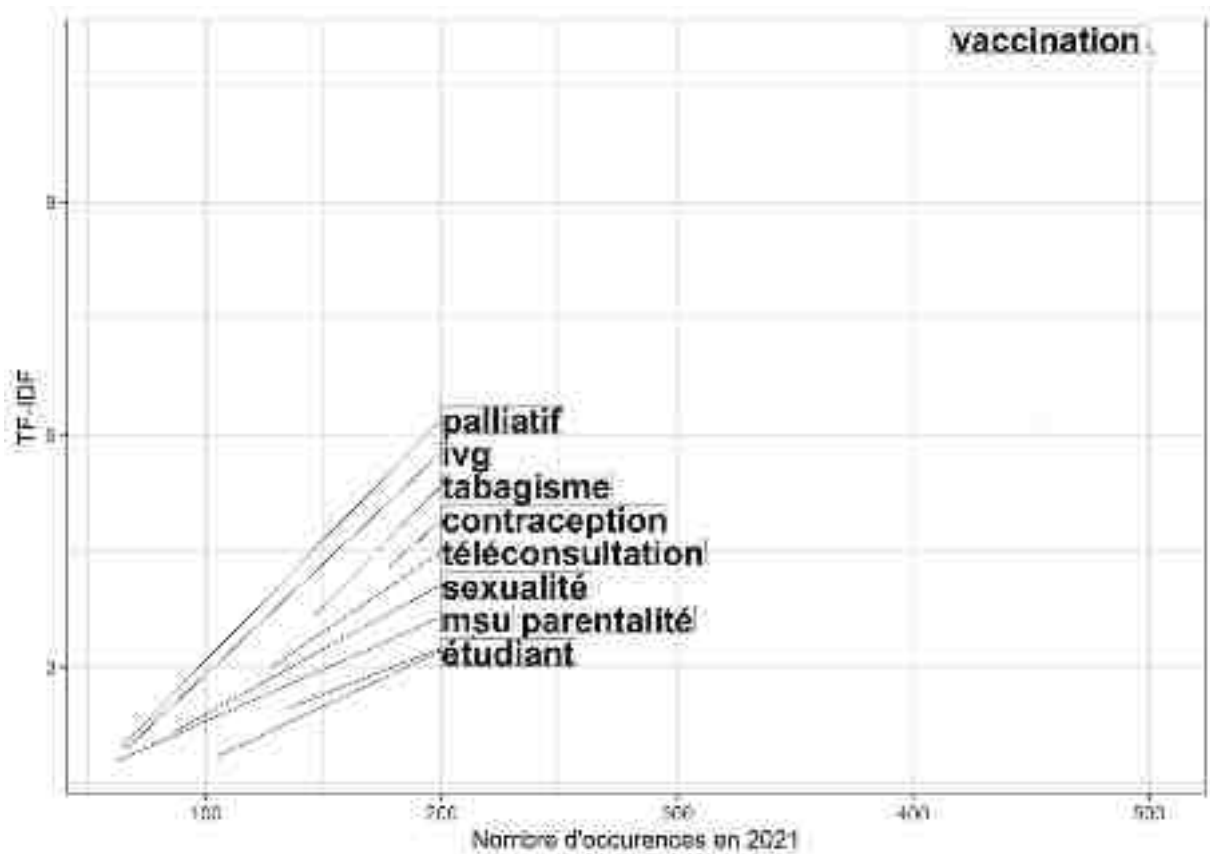


Figure 19 : Graphique du TF-IDF en fonction du nombre d'occurrences en 2021

Comme nous l'observons sur la Figure 19, le terme « vaccination » est de loin le plus représenté, suivi des termes « contraception », « tabagisme », « téléconsultation », puis « parentalité ». Les termes suivants sont « étudiant », « sexualité », « ivg », « palliatif », « msu » et « ipp ».

3.2.5 Analyse des bigrammes

3.2.5.1 Décompte de tous les bigrammes les plus fréquents par année

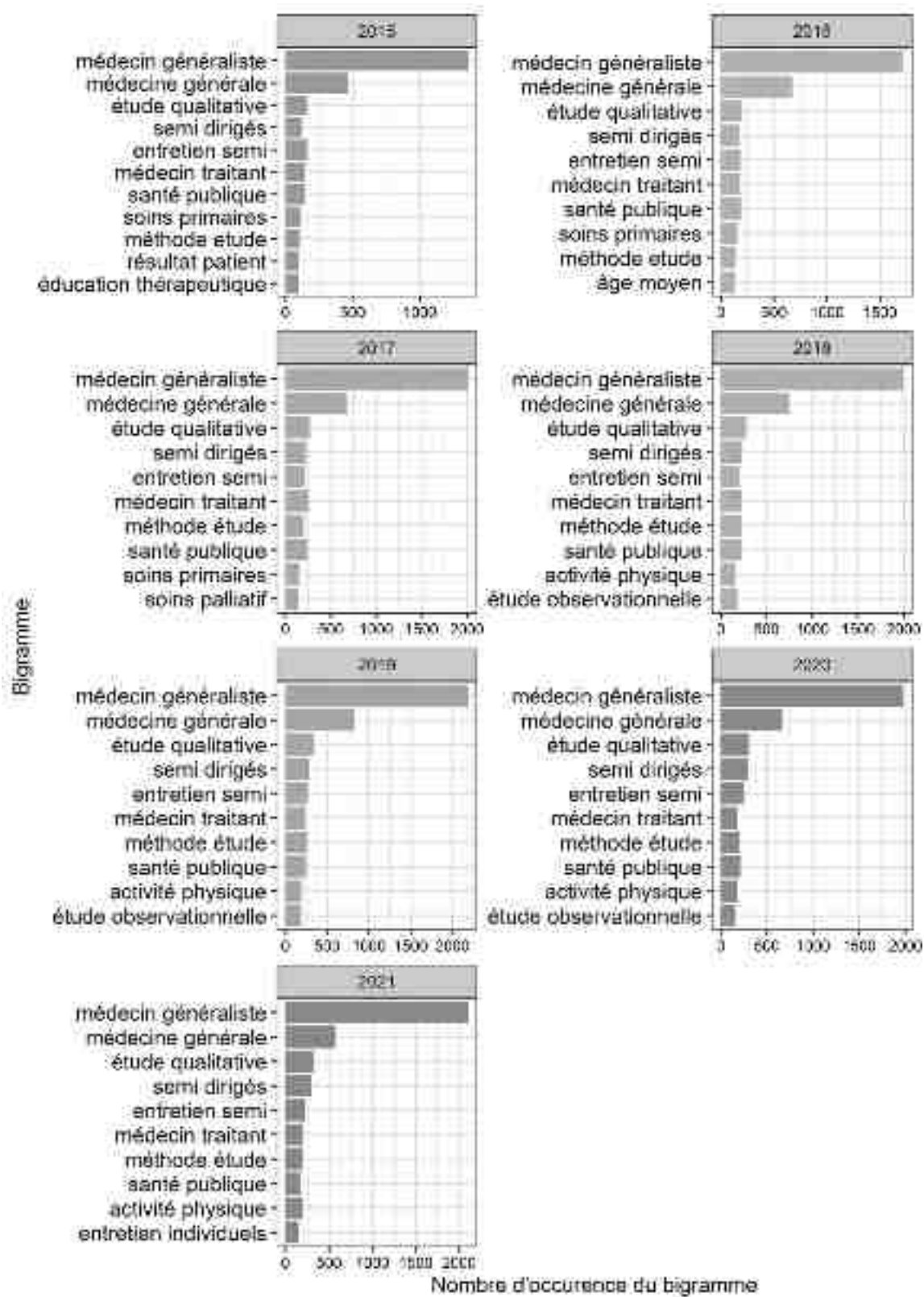


Figure 20 : Décompte des bigrammes les plus fréquents de 2015 à 2021

Le décompte des bigrammes les plus fréquents sur la Figure 20 montre qu'ils varient peu en fonction des années. On retrouve ici quelques thématiques déjà identifiées telles que « activité physique » ou « soins palliatifs ». Le reste des bigrammes concernent des expressions générales ou faisant référence à la méthodologie.

3.2.5.2 Fréquence d'association des bigrammes entre eux en 2021

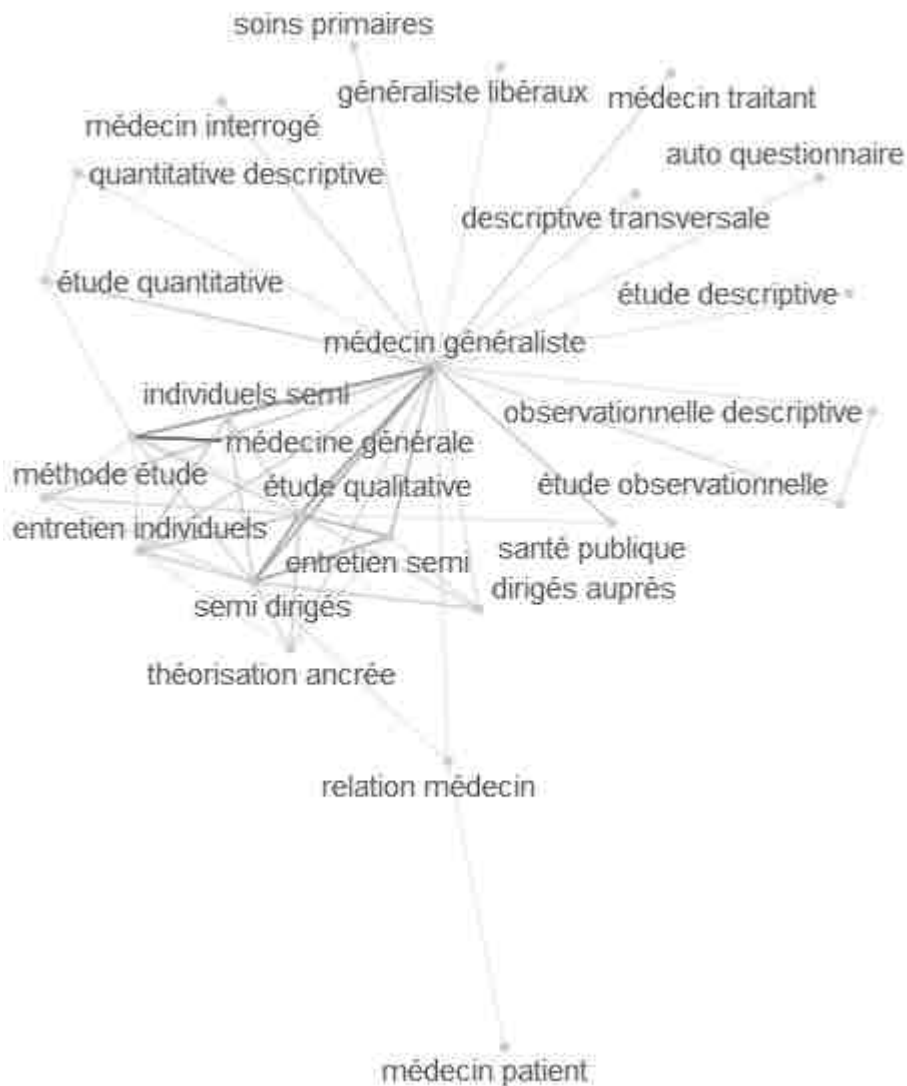


Figure 21 : Fréquence d'association des bigrammes entre eux sans tri des termes

La fréquence d'association des bigrammes entre eux montre elle aussi des expressions générales ou faisant référence à la méthodologie (Figure 21).

3.2.5.3 Corrélation des bigrammes entre eux avec tri des termes



Figure 22 : Corrélation de plus de 25 % des bigrammes apparaissant plus de 10 fois dans les résumés de thèses en 2021 avec tri des termes

La Figure 22 ci-dessus nous permet d'observer des corrélations entre certains bigrammes. Les nouvelles thématiques qui se dessinent ici concernent le diabète et la santé mentale.

3.2.5.4 Fréquence des bigrammes et fréquence de document inverse



Figure 23 : Nuage de mots des 10 plus grands TF-IDF des bigrammes de 2021

En utilisant le TF-IDF nous identifions quatre thématiques supplémentaires (Figure 23) :

- « *post partum* »,
- « *diabète gestationnel* »,
- « *fibrose hépatique* »,
- « *chirurgie bariatrique* ».

3.3 Fouille de texte par apprentissage automatique

3.3.1 Échantillonnage de Gibbs sur l'année 2021

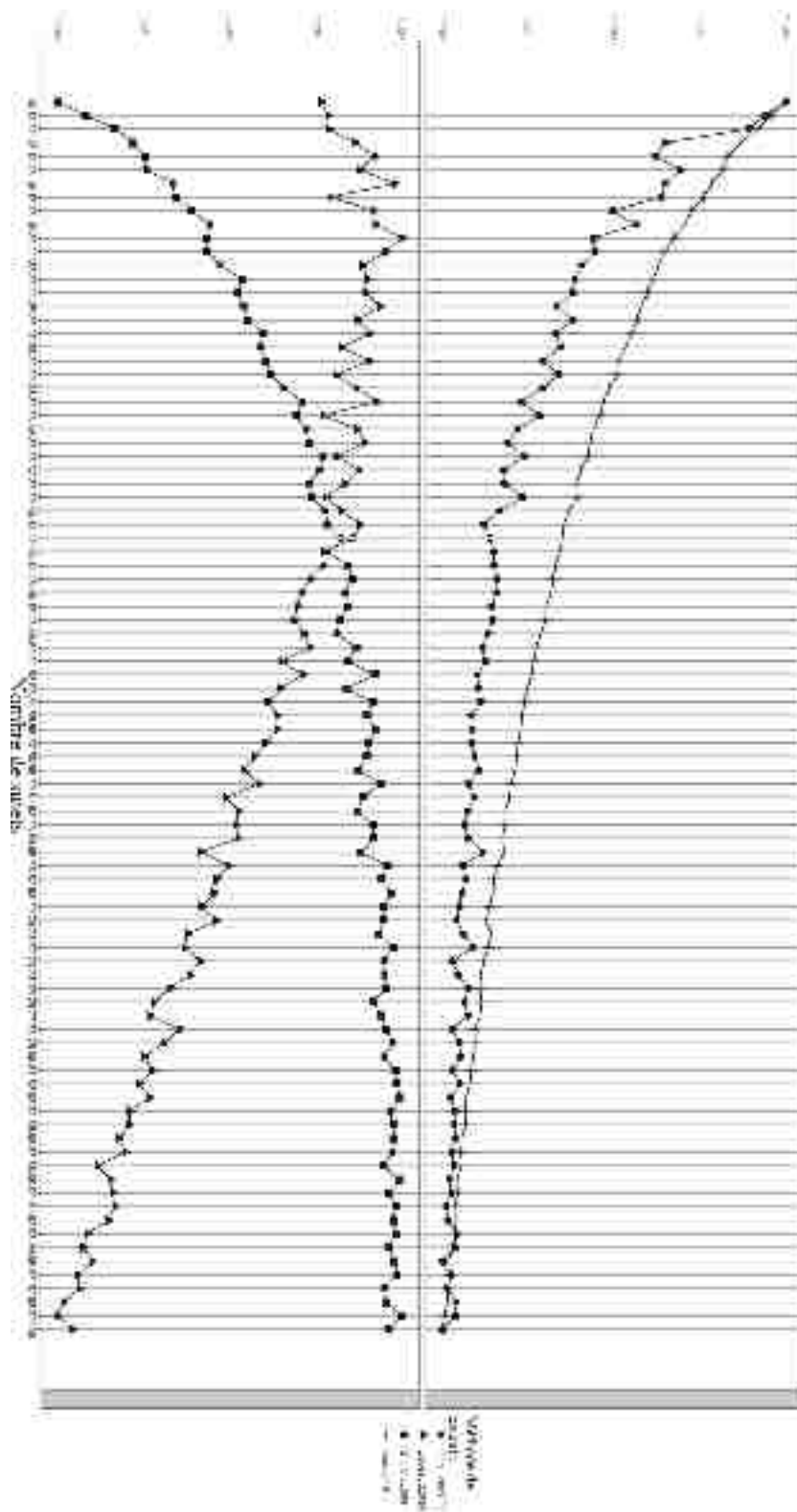


Figure 24 : Échantillonnage de Gibbs de 10 à 100 sujets en 2021

De loin l'opération la plus gourmande en ressource de calcul, l'échantillonnage de Gibbs nous permet d'approximer statistiquement le nombre de sujets compris dans le corpus des résumés de thèse de 2021. Sur la Figure 24 nous observons la représentation graphique du résultat du calcul selon les quatre méthodes. Les méthodes de Deveaud2014 et Griffiths2004 (49,50) tendent vers 1 lorsque le nombre de sujets est cohérent et les méthodes de Caojuan2009 et Arun2010 (47,48) tendent vers 0 lorsque le nombre de sujets est cohérent. En choisissant un point sur la courbe ou les méthodes de Deveaud2014 et Griffiths2004 sont maximum et les méthodes de Caojuan2009 et Arun2010 sont minimum. Nous pouvons estimer que le nombre optimal de sujets à rechercher est de 42.

3.3.2 Allocation latente de Dirichlet

3.3.2.1 Résultats de l'Allocation latente de Dirichlet

Tableau 3 : Probabilités de présence d'un sujet dans un résumé de thèse obtenue par allocation latente de Dirichlet et interprétations

Probabilité	Les 8 premiers termes racinisés classés par allocation latente de Dirichlet	Interprétation
5,9	entretien médecin qualit sem général dirig réalis aupres	Étude qualitative
5,2	médecin général questionnair pratiq réalis étud descript quantit	Étude quantitative
4,2	âge facteur popul plus étud risqu chez prévalent	Facteurs de risques
3,7	médecin format général manqu frein pratiq connaiss interrog	Formation des médecins généralistes
3,4	charg pris médecin général diagnostic patholog consult examen	Pratiques en médecine générale
3,2	group signif mois entre étud apres intervent inclus	Méthodologie
3,2	patient hospitalis servic hôpital hospitali rétrospect étud urgenc	Patients hospitalisés
3,2	patient médecin général consult relat médical inform ressent	Relation médecin-patient
3	connaiss questionnair inform répondr étud avoir question répond	Méthodologie
2,9	professionnel sant commun soin amélior coordin protocol organis	Réseaux de soins
2,7	revu articl littérature don étud recherch systémat bas	Revue de la littérature
2,6	sant prévent professionnel impact import malad travail fait	Santé au travail
2,5	médecin install général exercic médical libéral plus rural	Installation et exercice de la médecine générale en milieu rural
2,4	social soign question particip vécu médical représent expérient	Indéterminé
2,4	covid confin pandem épidem sanitair cris mar premi	COVID
2,4	scor évalú test échograph analys entre cliniqu mesur	Outils diagnostiques
2,3	fich inform outil sit expert utilis internet valid	Méthodologie
2,3	urgenc soin consult médical patient demand recours régul	Recours aux urgences
2,3	utilis téléconsult pratiq outil médecin consult patient télémedecin	Télémedecine
2,2	cas patient diagnostic infect présent cliniqu chez grav	Étude de cas
2,2	soin patient sant parcour social pass acces sein	Accès aux soins
2,2	travail afin mis thes plac anné part action	Méthodologie
2,2	patient risqu trait mortal chez associ facteur cardiaqu	Facteurs de risques
2,1	intern format stag msu étudi compétent enseign univers	Formation des internes
1,9	dépistag violenc femm canc victim réalis conjugal systémat	Violences conjugales
1,9	physiqu activ patient pratiq programm sport particip apa	Activité physique
1,8	prescript médic antibiothérap indiqu consomm ipp cour étud	Pratiques de prescription
1,8	troubl sommeil dépress psychiatr symptôm diagnostic patholog présent	Troubles du sommeil et psychiatriques
1,8	âgé person gériatr ehpad fragil âge résident chut	Personne âgé fragile
1,8	femm contracept gynécolog ivg utilis suiv diu sag	Suivi gynécologique et contraception chez la femme
1,8	douleur patient trait thérapeut utilis chroniqu antalq moyen	Prise en charge des douleurs chroniques
1,7	diabet patient suiv typ éduc malad etp thérapeut	Diabète
1,7	enfant parental pédiatr suiv âge écran chez conseil	Suivi pédiatrique et relation aux écrans chez les enfants
1,7	soin patient domicil palliat aid anticip fin direct	Soins à domicile et fin de vie
1,7	sexual adolescent consult jeun ist abord prévent éduc	Education sexuelle et prévention des IST
1,7	vaccin contr couvertur hpv campagn recommand gripp inform	Vaccination
1,6	recommand prévent âge appliqu pratiq vitamin primair cibl	Médecine préventive
1,5	étudi soign stress chez empath entre professionnel burnout	Syndrome d'épuisement professionnel chez les soignants et les étudiants
1,5	tabag consomm sevrage alcool addictolog arrêt aid cigaret	Sevrage alcool et tabac
1,4	femm grossess allait enceint nourrisson mer matern post	Grossesse et maternité
1,4	sport the lésion traumat and pratiq certificat cours	Sport
1	obes poid chirurg surpoid nutritionnel opératoir oper pert	Prise en charge du surpoids et de l'obésité

Ci-dessus le Tableau 3 représente les résultats de l'allocation latente de Dirichlet, avec les probabilités de présence d'un sujet dans la première colonne, arrondies à deux chiffres significatifs. Dans la deuxième colonne, nous retrouvons les huit premiers mots que l'algorithme a classé dans chaque catégorie. La troisième colonne renseigne leur interprétation, basée sur ces 8 mots. Nous observons que certains sujets, qui avaient été écartés des analyses précédentes, réapparaissent, notamment les sujets sur les types d'étude et la méthodologie. Certains d'entre eux avaient déjà été identifiés par les précédentes analyses. . Nous identifions également la mise en évidence d'autres sujets tels que :

- les facteurs de risques (2,2%),
- la formation des médecins généralistes (3,7%),
- les pratiques en médecine générale (3,4%),
- les relations médecin-patient (3,2%),
- la santé au travail (2,6%),
- les outils diagnostiques (2,4%),
- la formation des internes (2,1%),
- la médecine préventive (1,6%).

3.3.2.2 Identification de sujets via les résultats obtenus

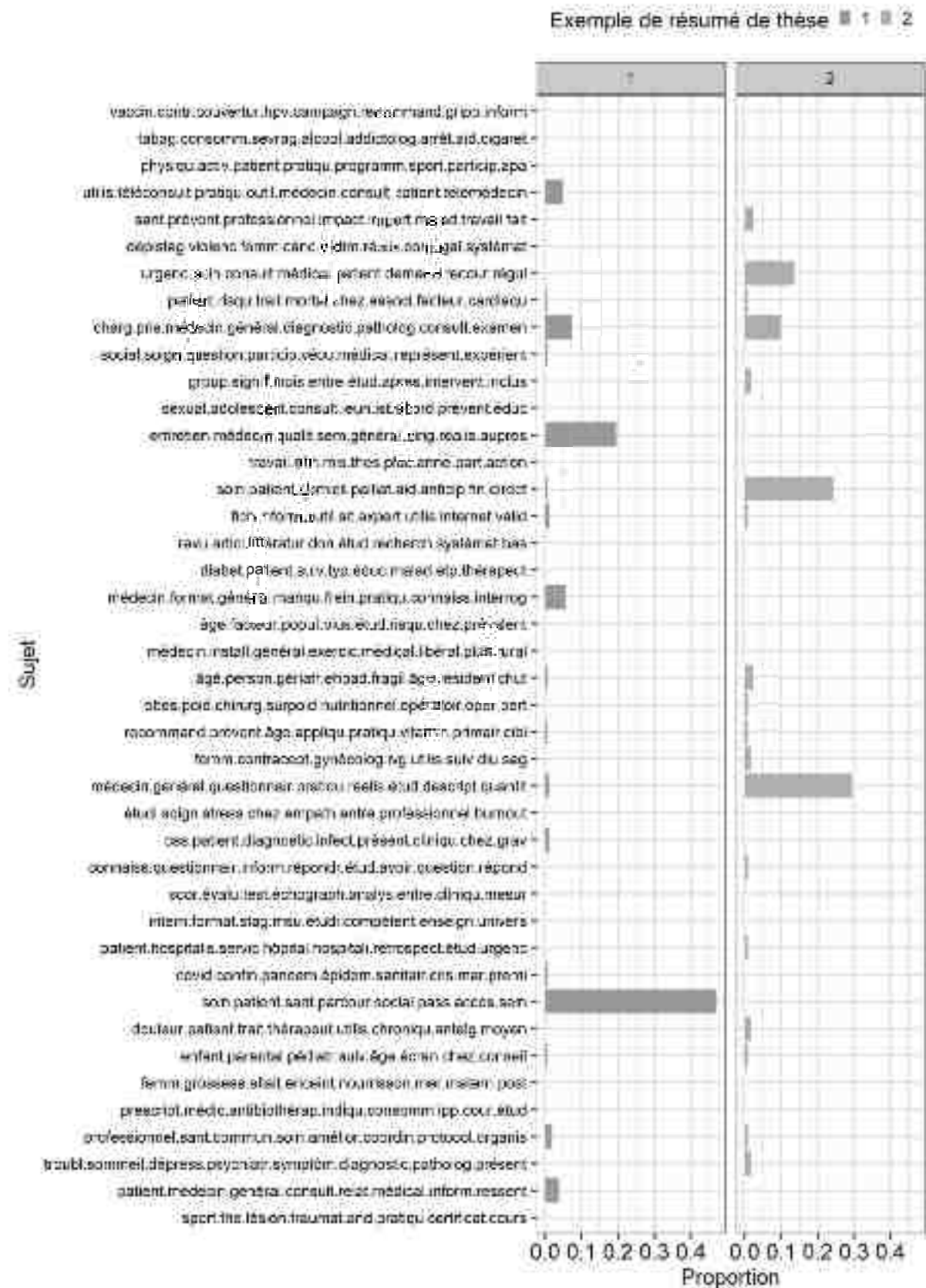


Figure 25 : Détermination du sujet de deux résumés de thèses du corpus

La Figure 25 illustre deux résumés de thèse que nous avons choisi au hasard sans en connaître le contenu. Nous avons ensuite utilisé les résultats de l'algorithme afin de détecter automatiquement les sujets de ces deux résumés. Nous observons que le document 1 va être en rapport avec l'accès aux soins et que le document 2 sera en rapport avec les soins à domicile et la fin de vie.

Le titre de la première thèse est « D'une PASS de ville aux soins primaires : expériences des patients pris en charge par le dispositif "PASS de Ville" de Médecins du Monde »(54). La deuxième thèse s'intitule quant à elle « État des lieux de la gestion à domicile par les médecins généralistes de la Somme des symptômes qui caractérisent les évolutions des pathologies qui affectent les patients palliatifs »(55). Nous observons donc que l'algorithme est capable de déterminer le sujet de ces thèses.

3.4 Évolution de certains sujets depuis 2015

Nous observons dans cette dernière analyse que la plupart des sujets sont similaires à ceux précédemment mis en évidence, donc nous pouvons en regrouper certains, afin de faciliter la lisibilité des résultats.

3.4.1 Résultats avec augmentation significative

3.4.1.1 Télémedecine

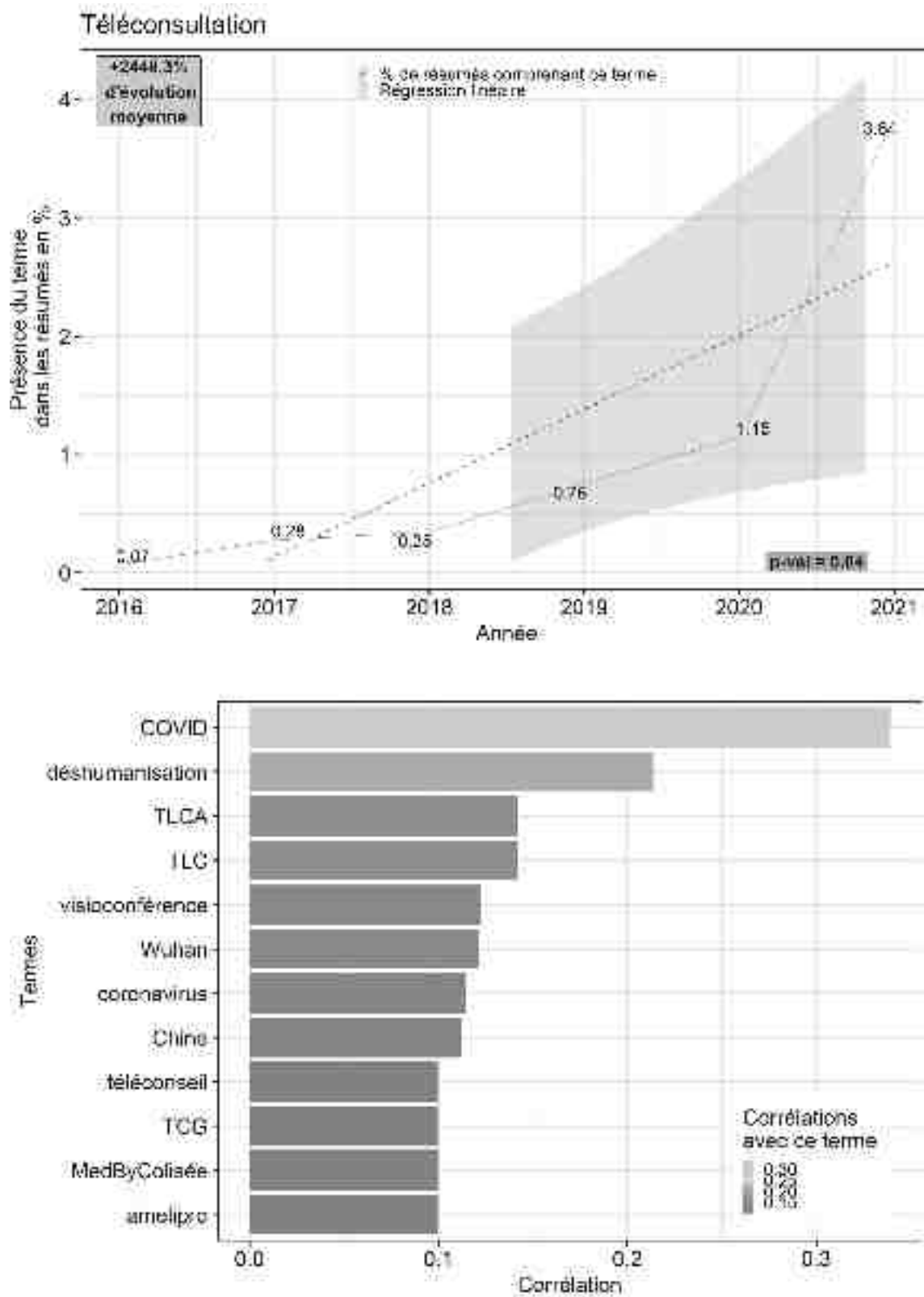


Figure 26 : Évolution et corrélations du terme téléconsultation sur la période 2016 à 2021

Tous les résultats suivants seront présentés sous le même format. Sur le graphique du haut nous observons deux courbes. Une courbe rouge qui représente la présence du terme étudié dans les résumés de thèse en pour cent en fonction des années. Nous observons aussi une courbe bleue en tiret discontinu qui représente la régression linéaire appliquée à l'évolution du nombre de thèses en rapport avec le sujet. La zone grisée autour de la régression linéaire représente l'intervalle de confiance. En haut à gauche se trouve l'évolution de la présence du terme étudié dans les résumés en pourcents, calculée grâce aux valeurs maximales et minimales de la régression linéaire modélisée. En bas à droite nous trouvons la valeur de la significativité p de la modélisation.

Sur le bas de la figure, nous retrouvons toujours les dix premiers mots ayant la corrélation la plus élevée avec le terme étudié.

Sur la Figure 26 nous observons l'évolution du terme téléconsultation, que l'on a choisi pour représenter l'évolution de la télémédecine. La période d'étude est de 2016 à 2021, car aucun résumé ne mentionnait ce terme en 2015. Nous pouvons y observer une très forte augmentation de +2448,3% ($p=0,04$)

Sur le graphique des corrélations, nous pouvons y observer que de nombreux termes sont en rapport avec la pandémie de coronavirus qui a débuté en 2019. Nous pouvons d'ailleurs observer une augmentation brusque de ce thème entre 2020 et 2021.

3.4.1.2 La relation aux écrans

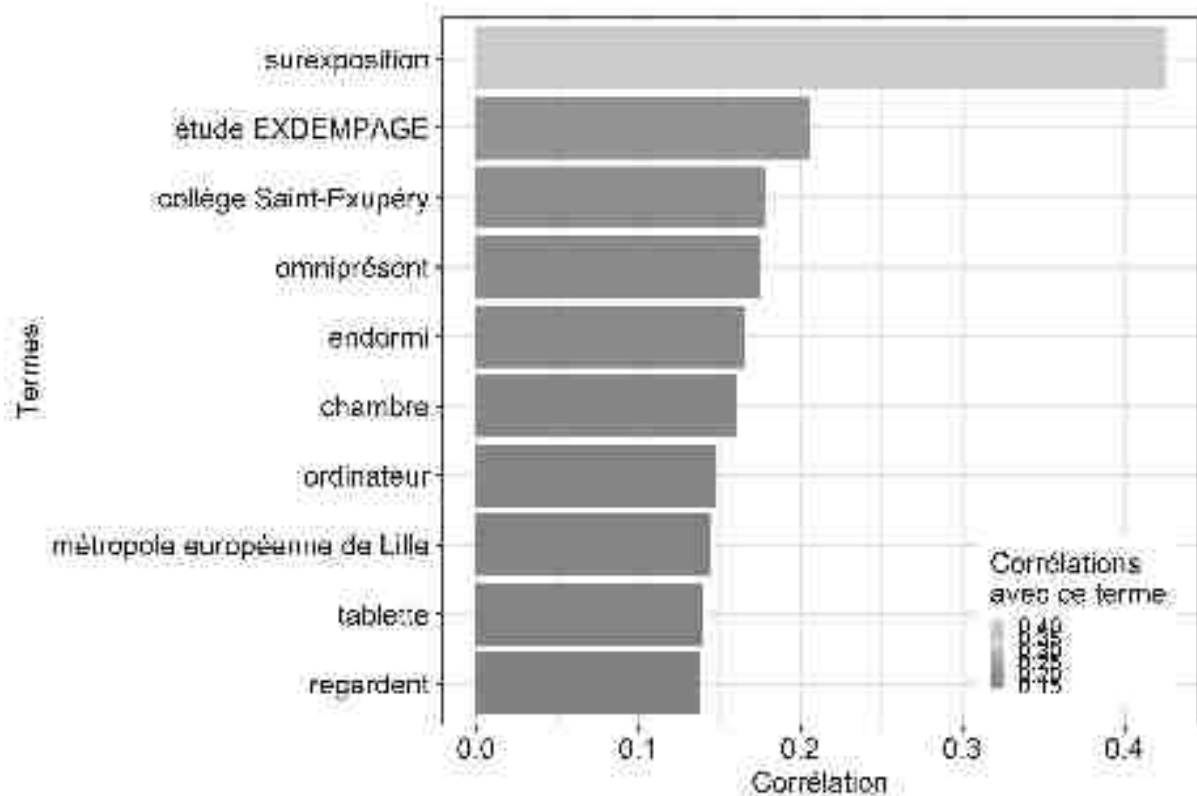
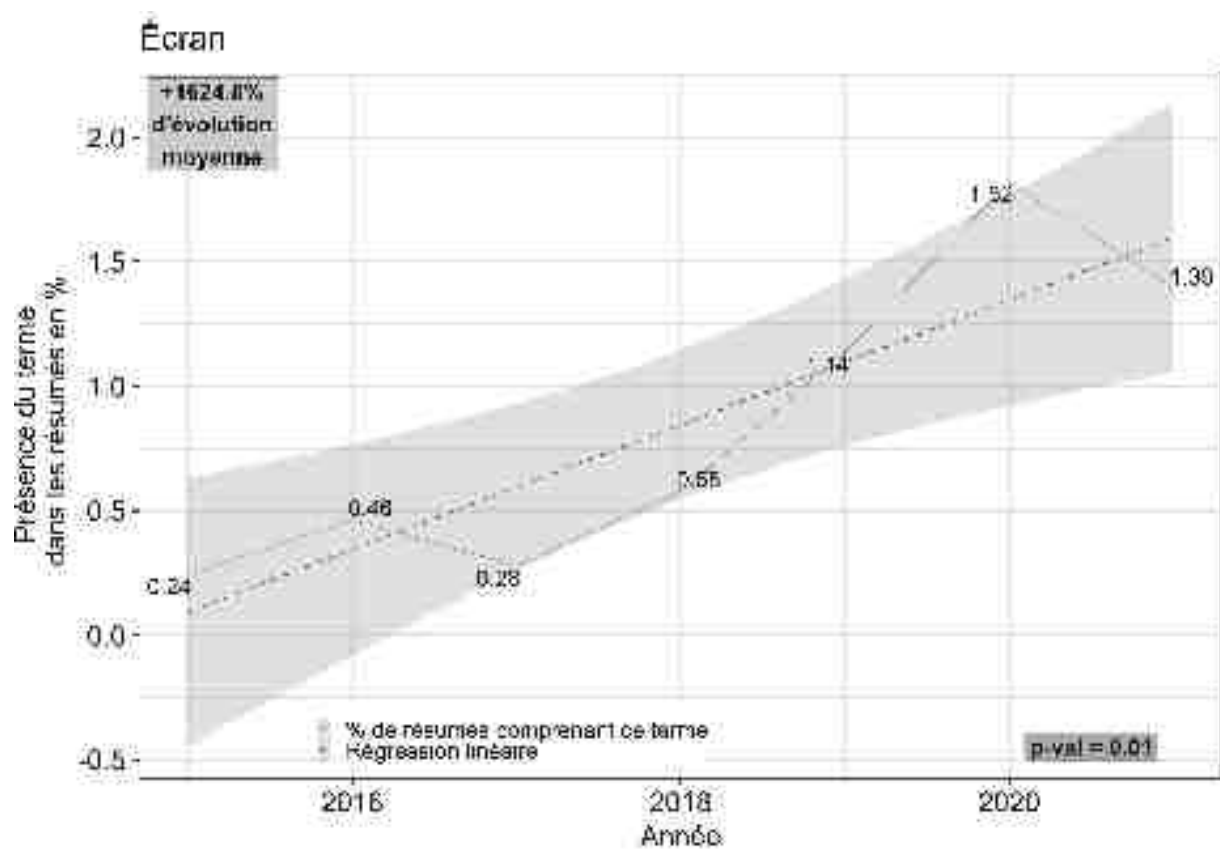


Figure 27 : Évolution et corrélations du terme écran sur la période 2015 à 2021

Sur la Figure 27 nous observons l'évolution du terme écran, que l'on a choisi pour représenter l'évolution de la relation aux écrans. Nous pouvons y observer une très forte augmentation de +1624,8% ($p=0,01$).

Sur le graphique des corrélations, nous pouvons y observer que de nombreux termes sont en rapport avec la notion de surexposition aux écrans, et que les termes suivants font référence à des études dont la population était des collégiens de la métropole de Lille (2).

Les autres termes sont en relation avec le sommeil. Ce terme reflète donc bien une recrudescence de l'intérêt des internes en médecine à propos de la question des écrans, notamment au niveau de l'enfance et de l'adolescence.

3.4.1.3 La COVID

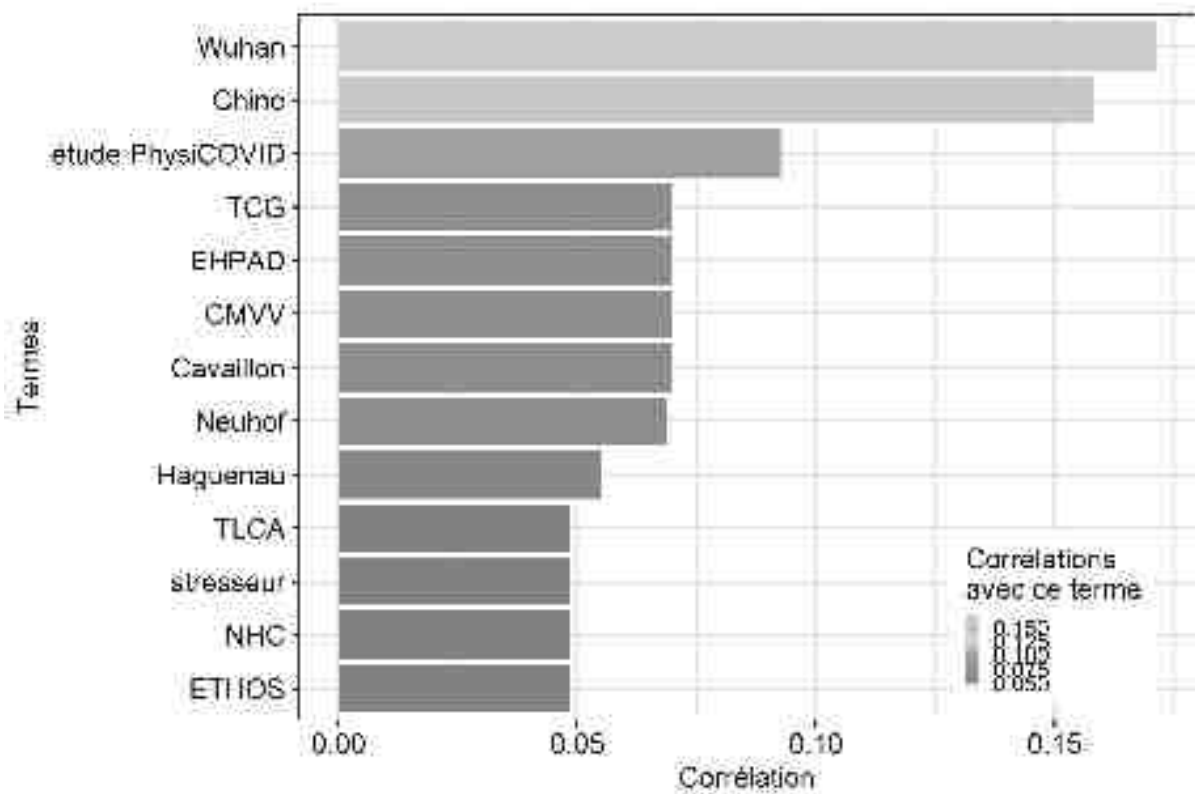
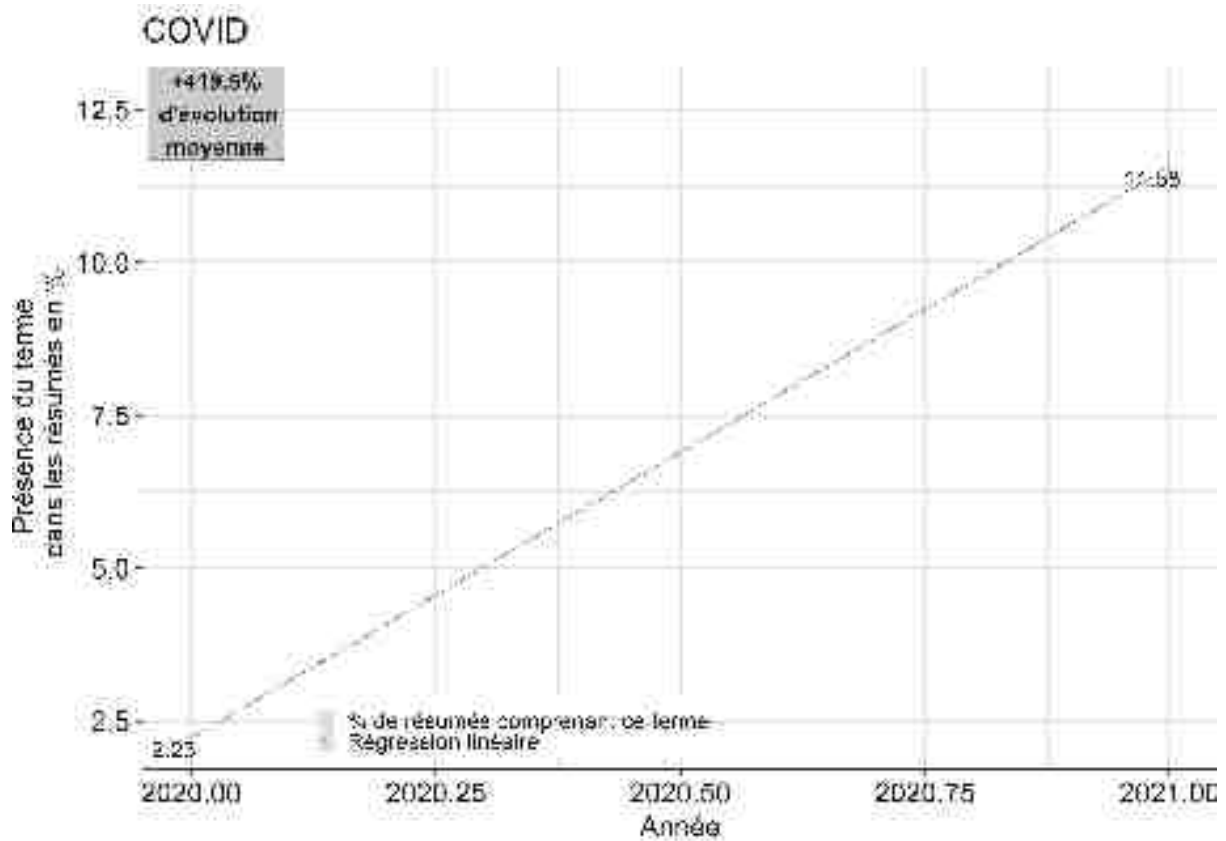


Figure 28 : Évolution et corrélations du terme COVID sur la période 2020 à 2021

Sur la Figure 28 nous observons l'évolution du terme COVID, que l'on a choisi pour représenter l'évolution de la pandémie de coronavirus. Nous pouvons y observer une très forte augmentation de +419,5%. Cependant, le terme COVID n'est présent que dans la période 2020 à 2021, il est donc pour l'instant hasardeux de présumer l'évolution future de ce sujet.

Sur le graphique des corrélations nous observons que des termes sont en relation avec l'origine supposée de la pandémie. De plus s'ajoutent des termes relatifs à l'étude physiCOVID (56), aux patients fragiles, notamment EHPAD et CMVV, ainsi que les termes TLCA et TCG relatifs à la télémédecine.

3.4.1.4 Violences

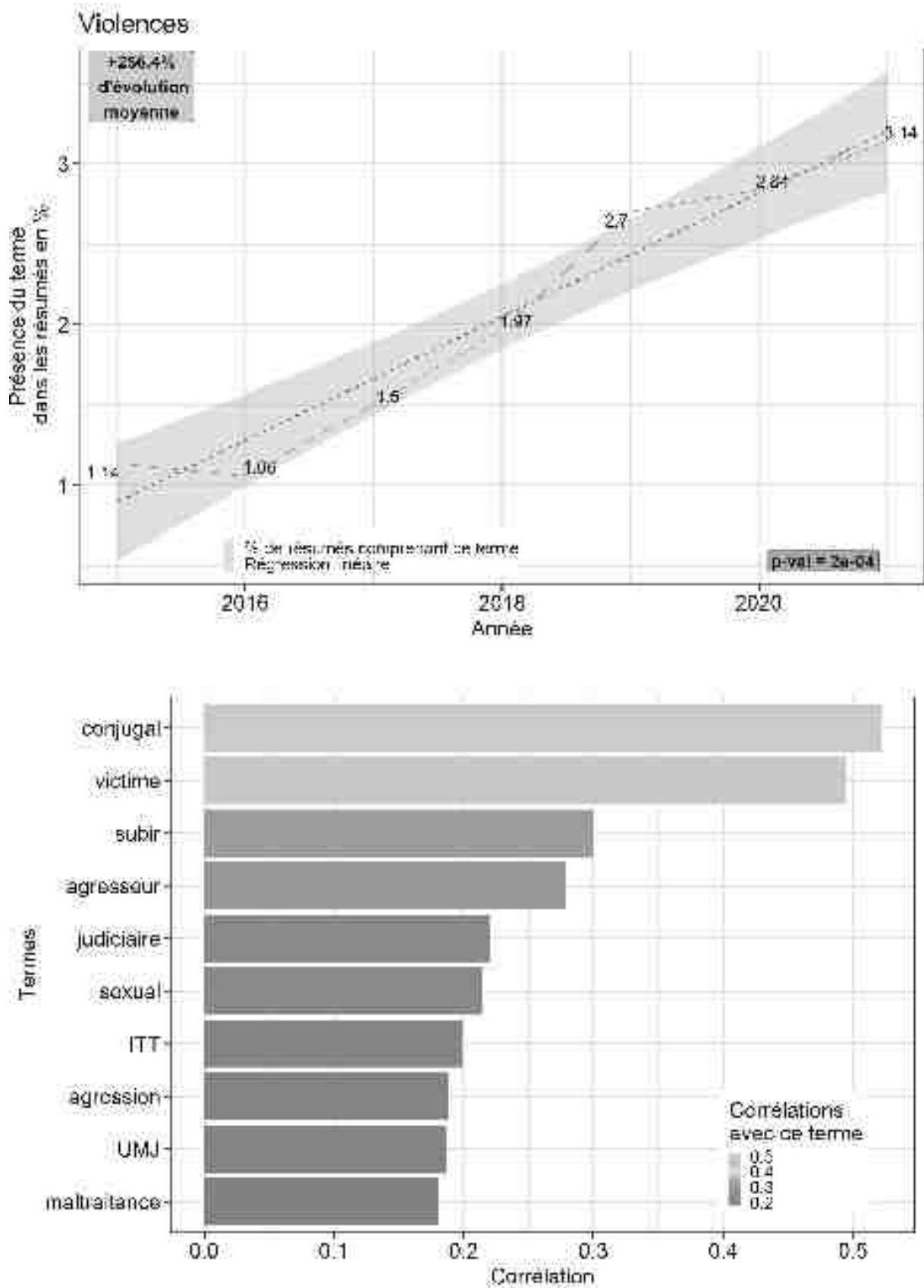


Figure 29 : Évolution et corrélations du terme violence sur la période 2015 à 2021

Sur la Figure 29 nous observons l'évolution du terme « violence », que l'on a choisi pour représenter l'évolution du thème des violences. Nous pouvons y observer une très forte augmentation régulière de +256,4% ($p=0,0002$) avec une significative importante.

3.4.1.5 Le parcours de soin

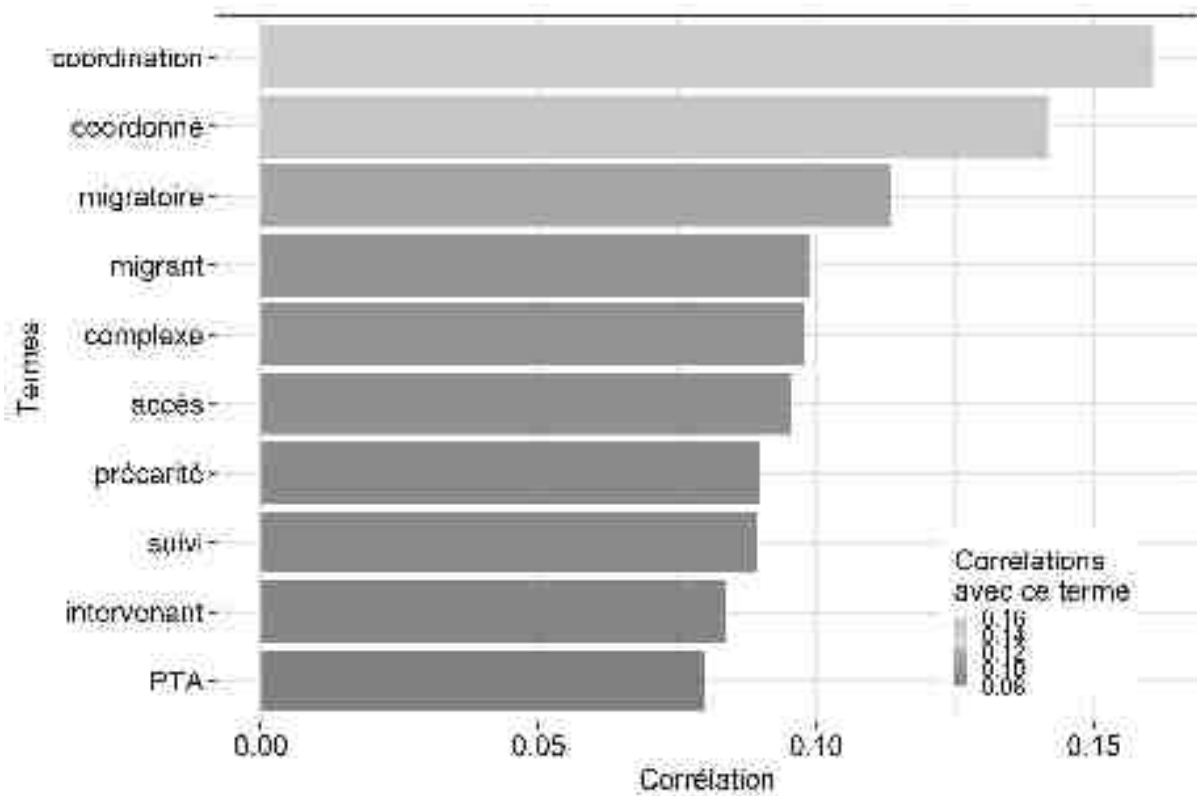
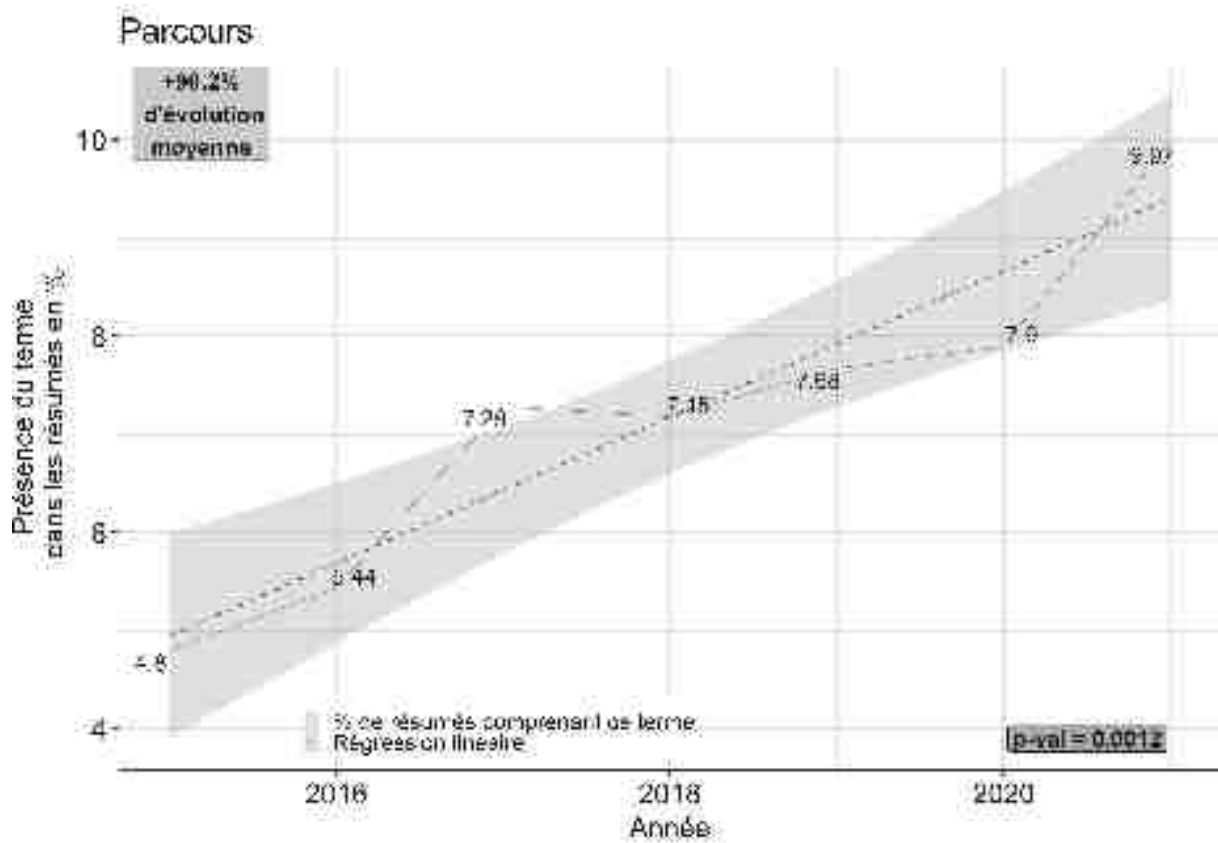


Figure 30 : Évolution et corrélations du terme parcours sur la période 2015 à 2021

Sur la Figure 30 : Évolution et corrélations du terme parcours sur la période 2015 à 2021 nous observons l'évolution du terme parcours, que l'on a sélectionné pour représenter l'évolution de la thématique du parcours de soins. Nous pouvons y observer une forte augmentation de +90,2% ($p=0,0012$) avec une significative importante.

Les parcours dans les résumés de thèses sont de plus en plus présents. Le parcours de soins coordonnés a été créé par la loi du 13 août 2004 (57). Il vise à assurer à chaque usager un suivi médical coordonné, une tenue rigoureuse du dossier médical et une prévention personnalisée.(58)

Ce thème est de plus en plus traité dans les thèses de médecine générale, avec la problématique de l'accès aux soins notamment par les migrants (59).

3.4.1.6 L'aspect Social

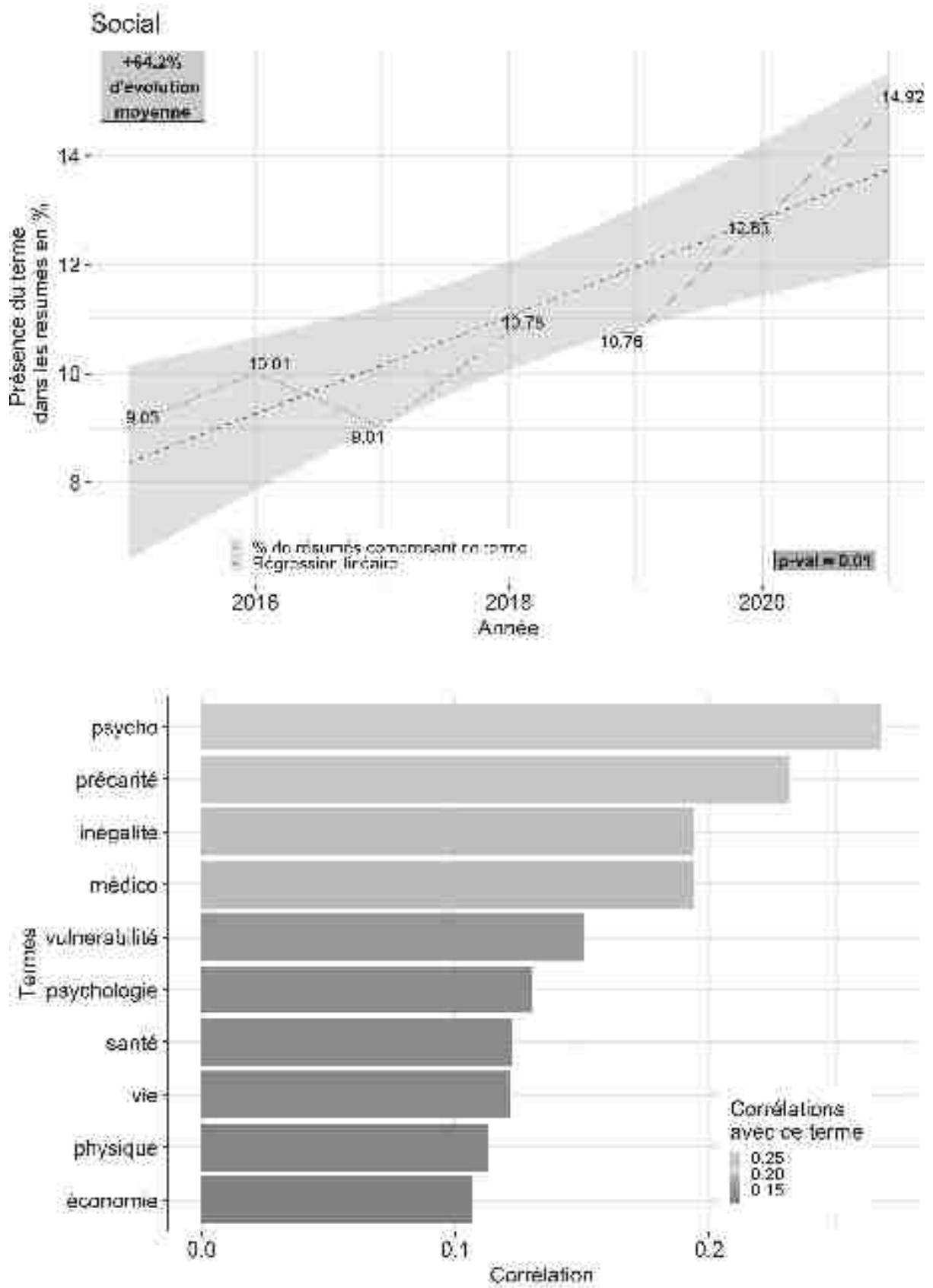


Figure 31 : Évolution et corrélations du terme social sur la période 2015 à 2021

Sur la Figure 31, nous observons le développement du terme social, qui a été choisi pour refléter le développement des thématiques sociales en médecine. Il y a une augmentation significative de +64,2% ($p=0,01$).

L'intérêt des internes de médecine générale est croissant en ce qui concerne les inégalités sociales en médecine. Ce sujet, le corps médical en a conscience depuis déjà plusieurs siècles. En 1700, Ramazzini (1633-1714) citait déjà dans son essai (60), les pathologies dont pouvaient souffrir certaines classes socio-professionnelles : « *La douleur des reins est encore une maladie particulière aux ouvriers assis. Ainsi Plaute a dit « Les lombes deviennent douloureuse lorsque l'on reste assis trop longtemps. » »* (Annexe 5).

Ces inégalités persistent encore aujourd'hui , avec par exemple une mortalité ouvrière plus importante avant la retraite que pour les cadres (61).

3.4.1.7 Vaccination

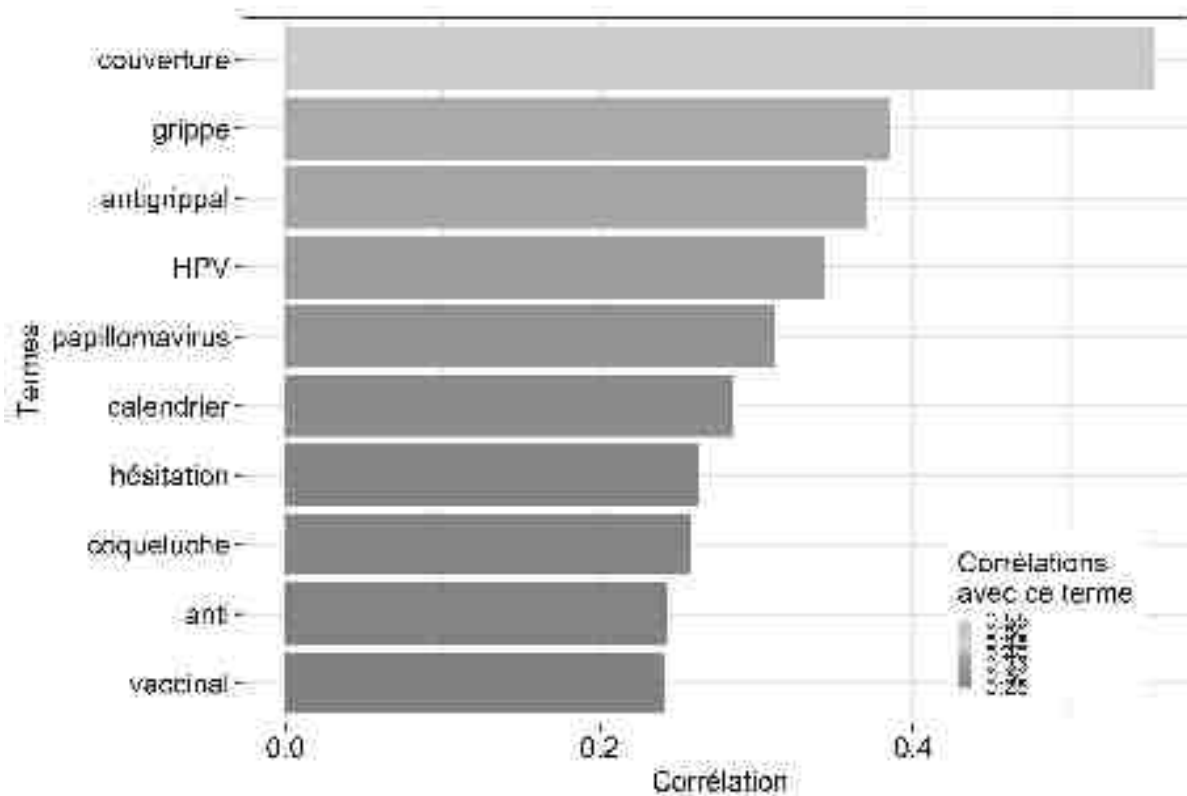
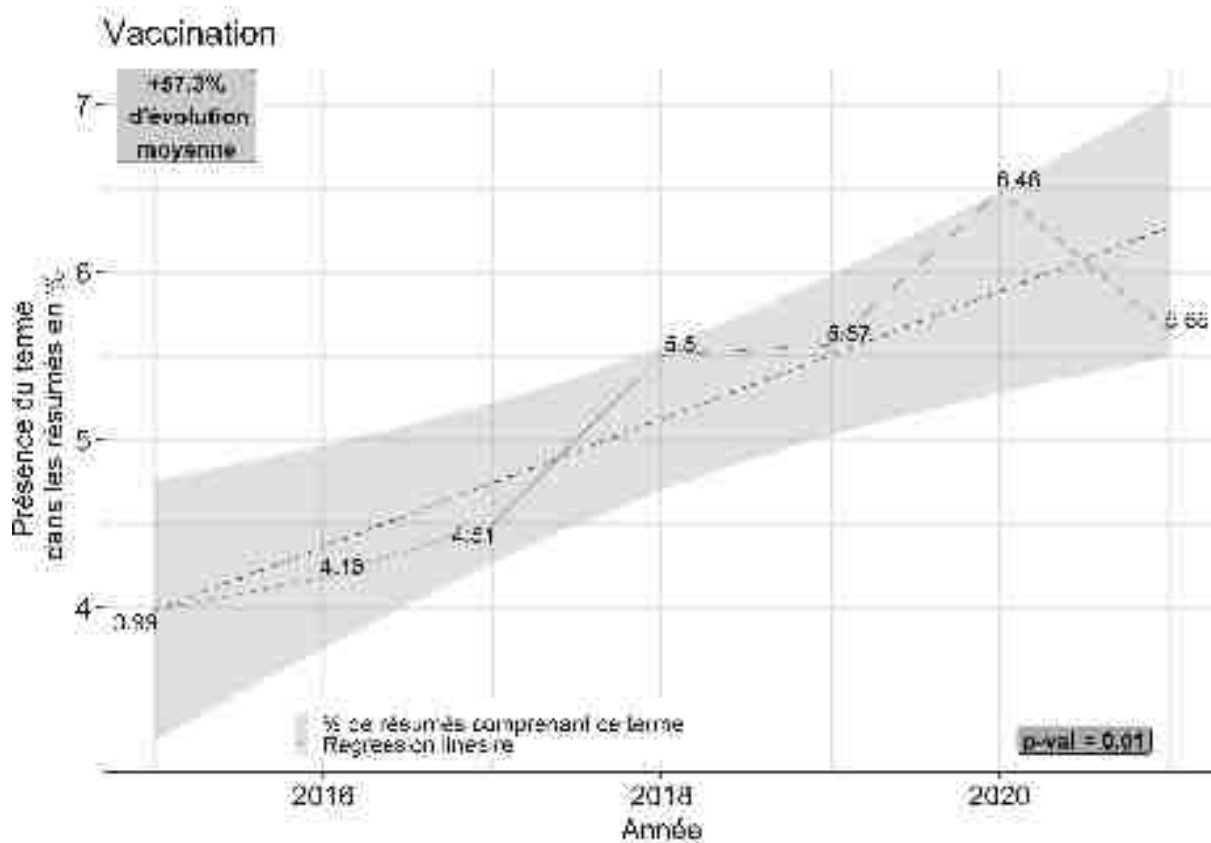


Figure 32 : Évolution et corrélations du terme vaccination sur la période 2015 à 2021

Sur la Figure 32, nous observons l'évolution du terme « vaccination », que l'on a choisi pour représenter l'évolution du thème de la vaccination. Nous pouvons y observer une forte augmentation de +57,3% ($p=0,01$)

La vaccination est un thème qui est en augmentation parmi les thèses de médecine générale. Cet intérêt est principalement centré sur la vaccination anti grippale et papillomavirus si l'on observe le graphique de corrélation. Le terme « hésitation » est aussi souvent corrélé à la vaccination. L'hésitation vaccinale est favorisée par des biais cognitifs sur la balance bénéfice-risque ainsi que par la propagation d'informations fausses ou invérifiables de manière personnalisée et automatisée sur les nouveaux moyens de communications comme par exemple les réseaux sociaux (62).

3.4.1.8 Activité sportive

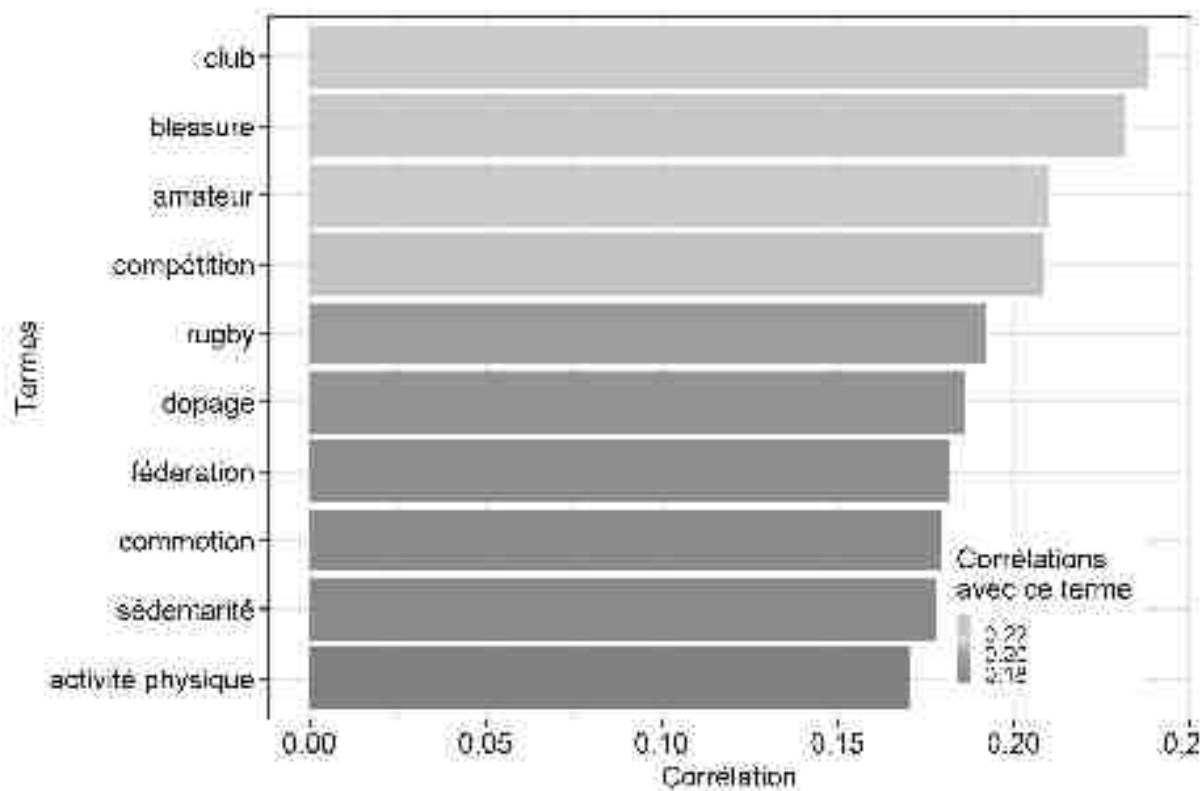
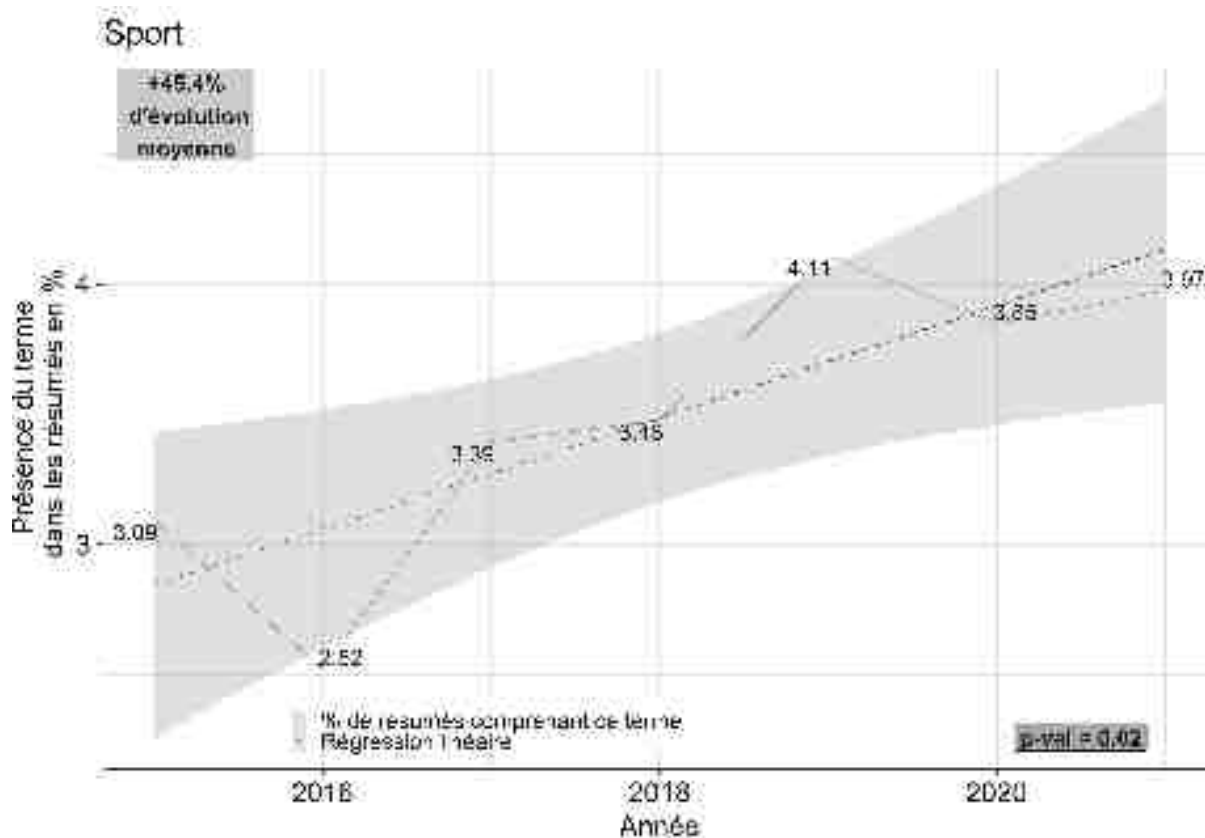


Figure 33 : Évolution et corrélations du terme sport sur la période 2015 à 2021

Sur la Figure 33, nous observons l'évolution du terme sport, que l'on a choisi pour représenter l'évolution de l'activité sportive. Nous pouvons y observer une augmentation de +45,4% ($p=0,02$)

Cette augmentation témoigne d'un intérêt croissant de la part des nouveaux médecins généralistes, en accord avec le programme national et nutrition (63), ainsi que l'augmentation de la pratique sportive et de l'activité physique dans la population générale (64). Nous notons que dans le graphique des corrélations, les risques de la pratique sportive y sont présents avec les termes blessure, dopage et commotion.

3.4.1.9 Allaitement

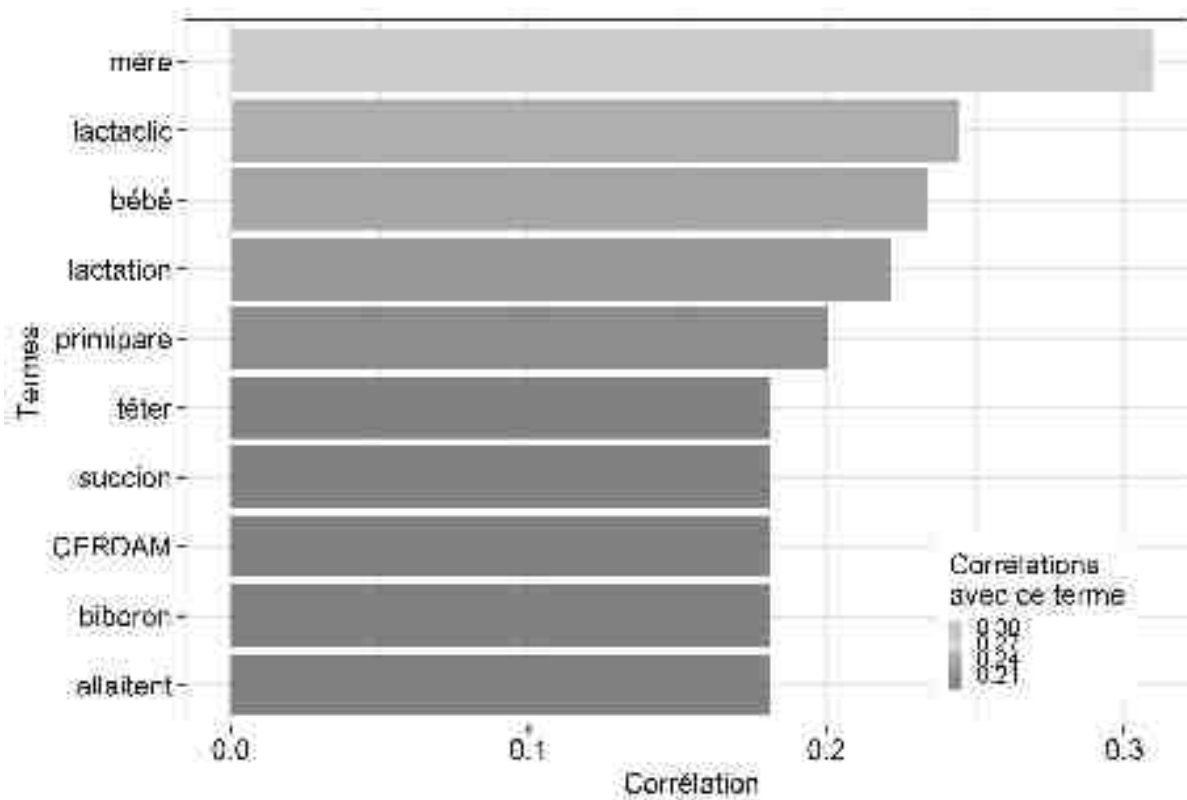
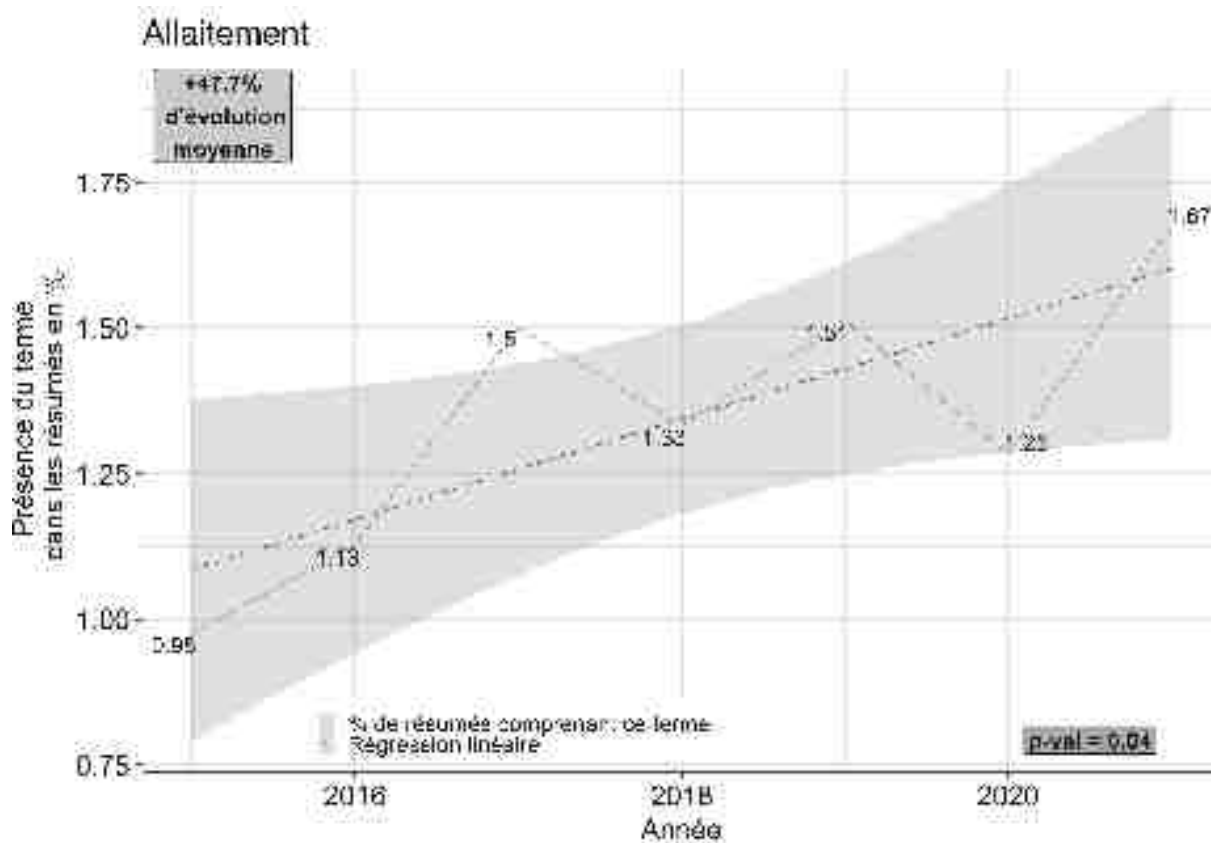


Figure 34 : Évolution et corrélations du terme allaitement sur la période 2015 à 2021

Sur la Figure 34, nous observons l'évolution du terme allaitement, représentatif du thème éponyme. Nous pouvons y observer une augmentation de +47,7% ($p=0,04$).

Sur le graphique des corrélations, nous pouvons y observer que les termes lactaclic et CERDAM font référence à des ressources utilisables pour le conseil à l'allaitement en médecine générale. Cela traduit un intérêt croissant de la part des internes de médecine générale à la promotion de l'allaitement, en accord avec les données actuelles de la science sur les bienfaits mère-enfant de l'allaitement (65).

3.4.1.10 Suivi pédiatrique en médecine générale

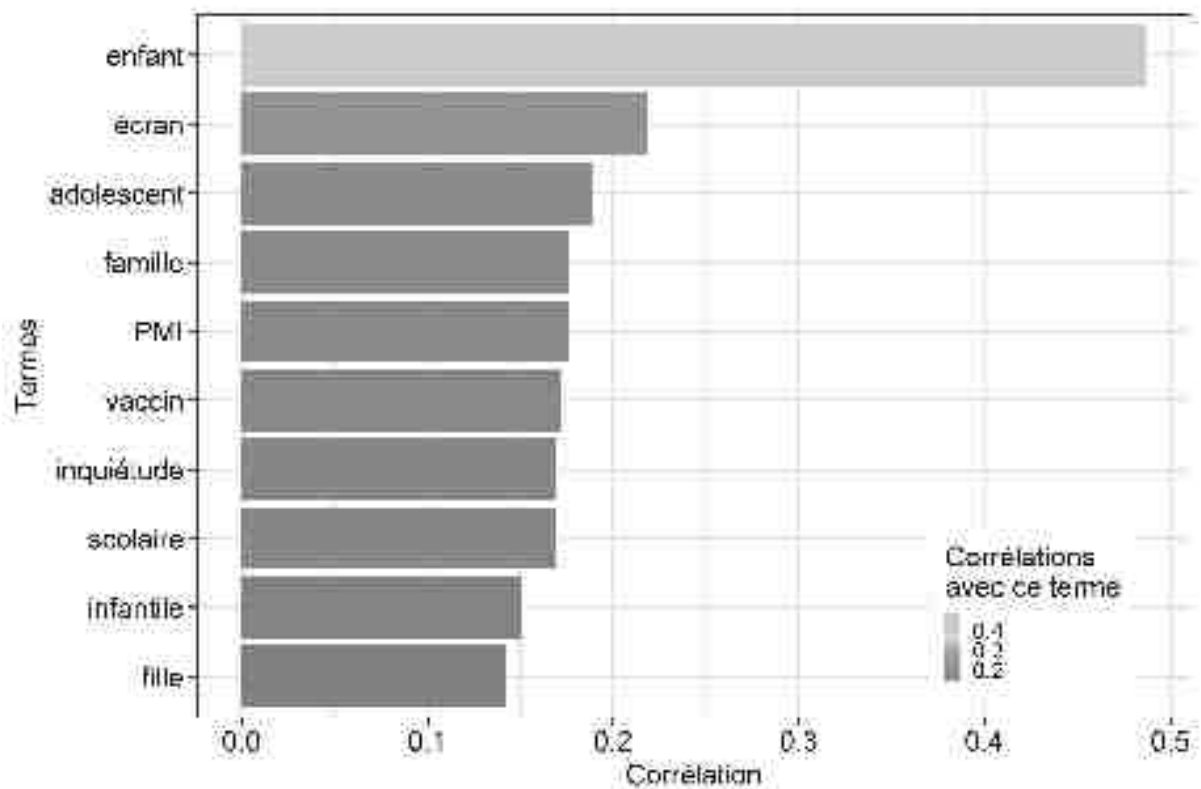
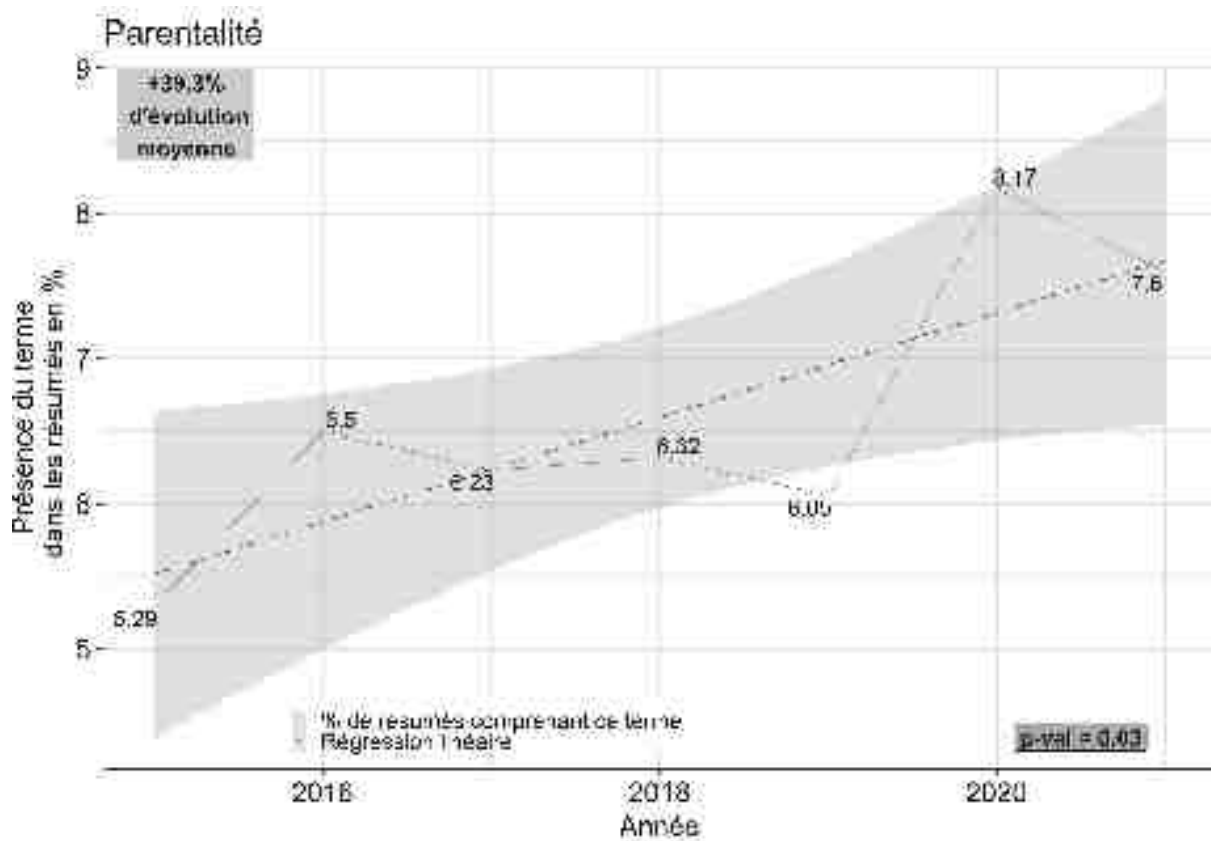


Figure 35 : Évolution et corrélations du terme parentalité sur la période 2015 à 2021

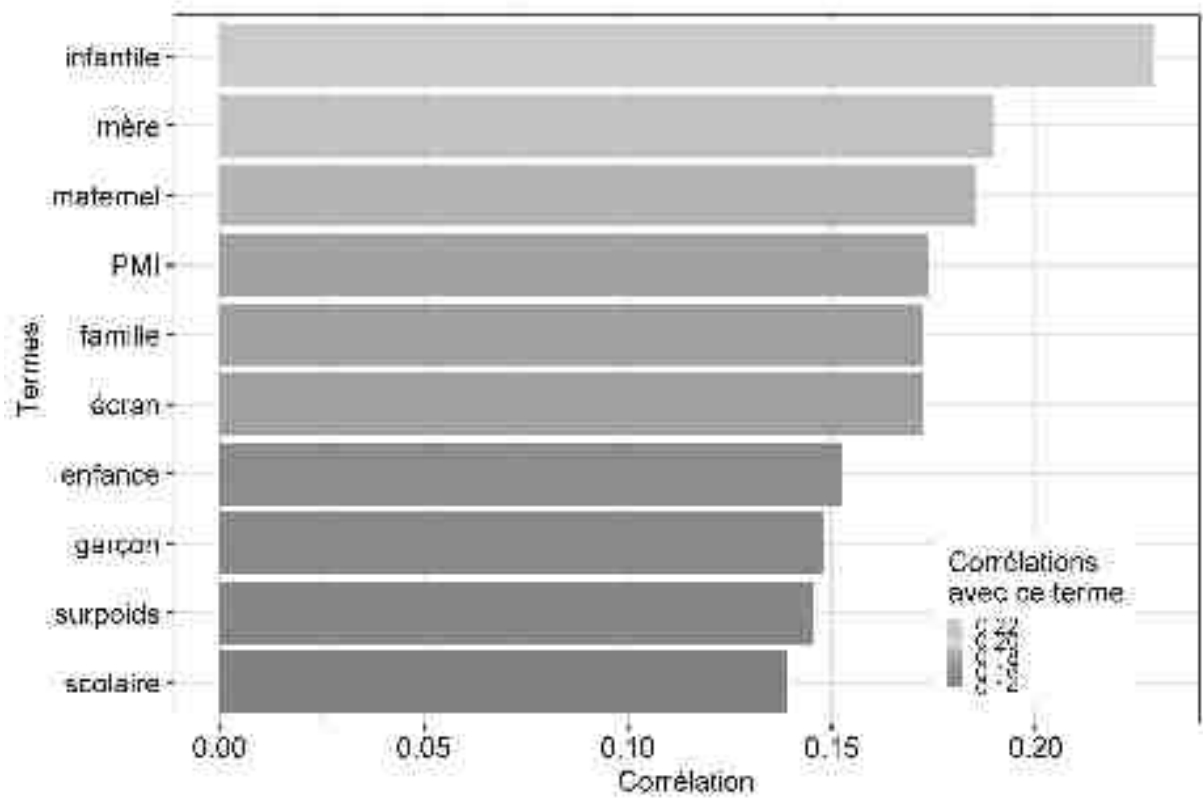
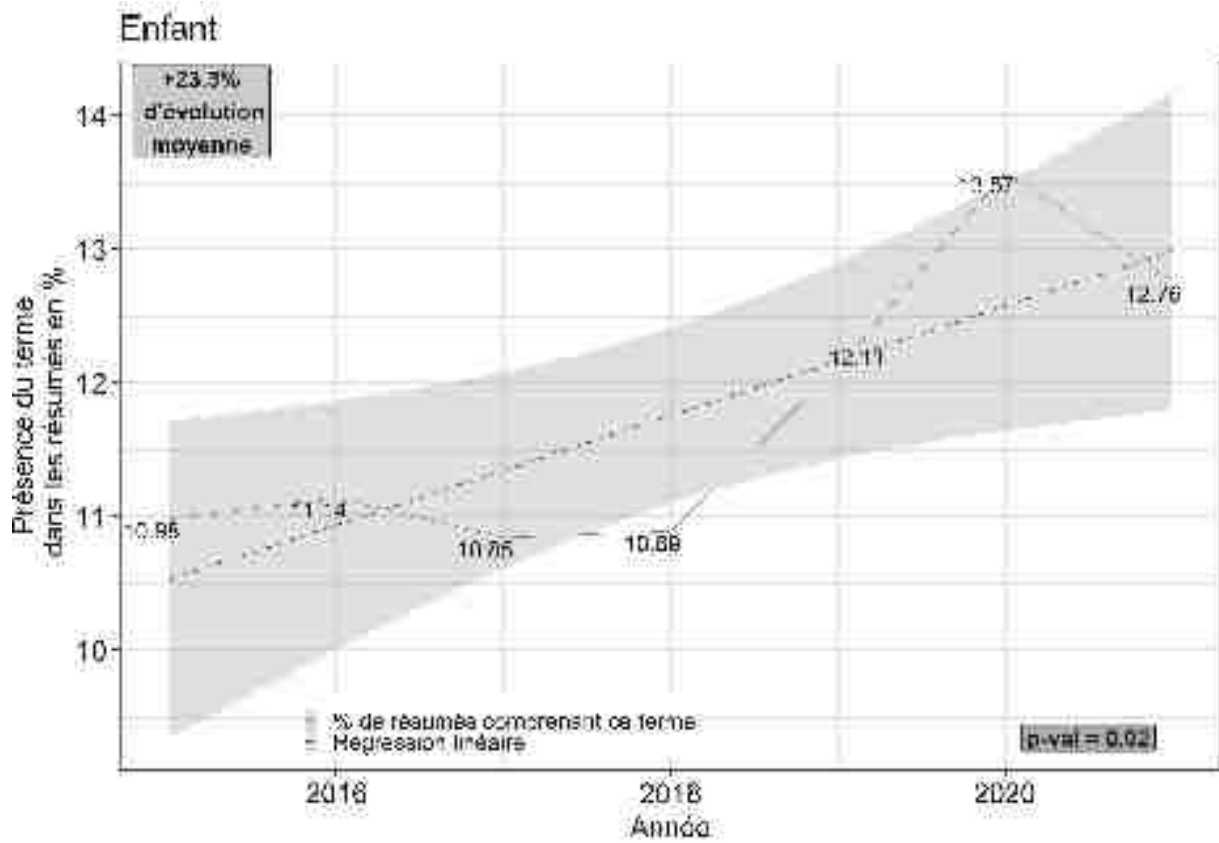


Figure 36 : Évolution et corrélations du terme enfant sur la période 2015 à 2021

Sur les Figure 35 et Figure 36, nous observons l'évolution des termes parentalité et enfant, représentatifs du suivi pédiatrique en médecine générale. Nous pouvons y observer une augmentation de respectivement +39,3% ($p=0,03$) et de +23,3% ($p=0,02$).

Nous retrouvons dans les corrélations des sujets abordés plus tôt comme la vaccination (Figure 32), et la relation aux écrans (Figure 27). La PMI pourrait aussi faire référence au sujet de la violence dans un contexte de maltraitance infantile (Figure 29 et Annexe 7 : Évolution et corrélations du terme maltraitance sur la période 2015 à 2021, page 151), cependant ce n'est sa pas seule mission.

Nous retrouverons aussi le sujet du surpoids qui est corrélé au suivi pédiatrique. Sur les Annexe 15 et Annexe 16 (pages 159-160) qui traitent du surpoids et de l'obésité nous observons la présence dans les diagrammes de corrélation de l'International Obesity Task Force (IOTF-30) qui est une définition du surpoids et de l'obésité chez l'enfant.

3.4.1.11 Suivi gynécologique et contraception chez la femme

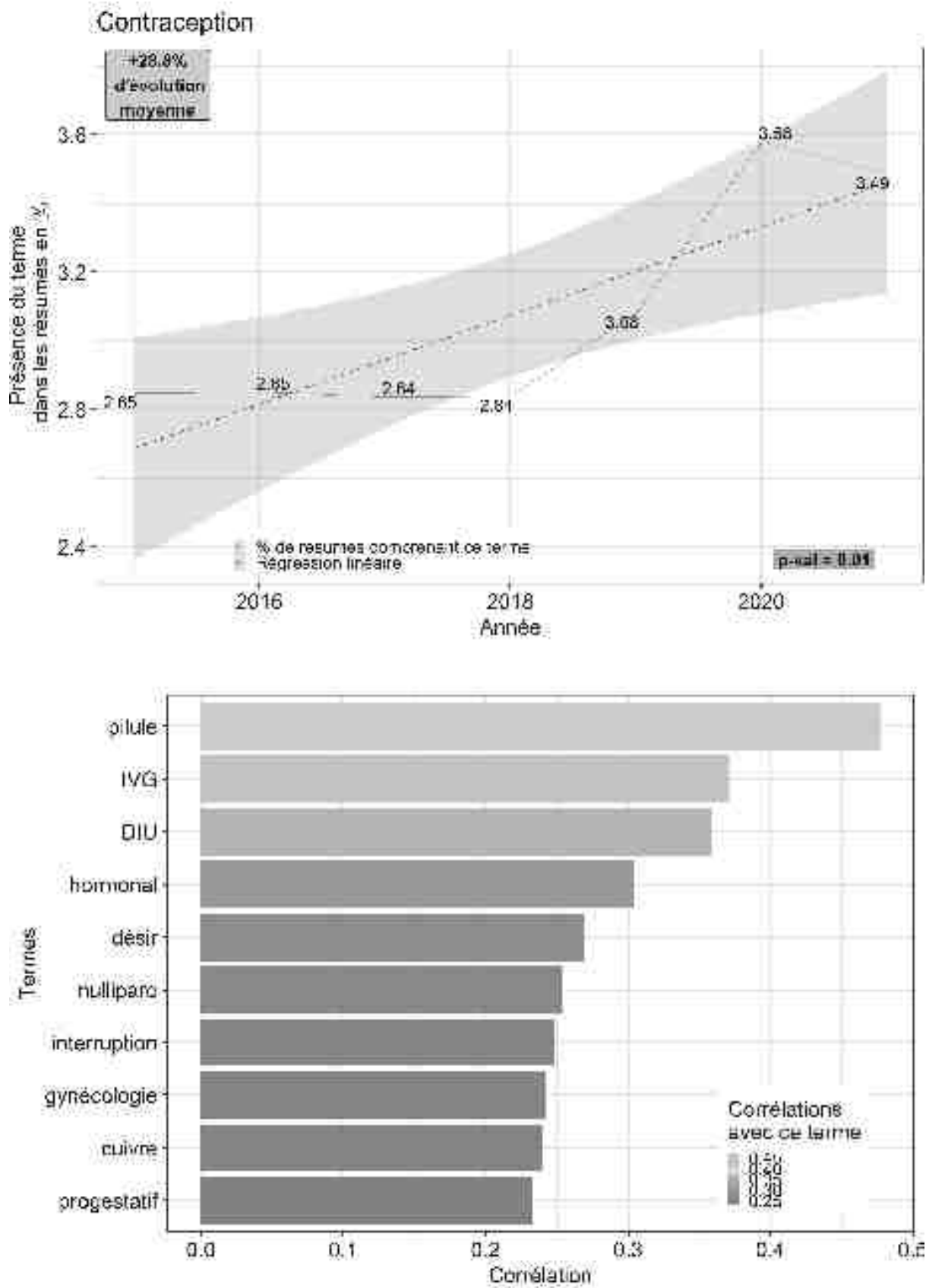


Figure 37 : Évolution et corrélations du terme contraception sur la période 2015 à 2021

Sur la Figure 37, nous observons l'évolution du terme contraception, en relation avec le thème du suivi gynécologique et contraception chez la femme. Nous pouvons y observer une augmentation de +28,8% ($p=0,01$).

Sur le graphique des corrélations, nous observons que le premier terme est pilule, et en effet la pilule reste le moyen de contraception le plus populaire (66).

3.4.1.12 Installation et exercice de la médecine générale en milieu rural

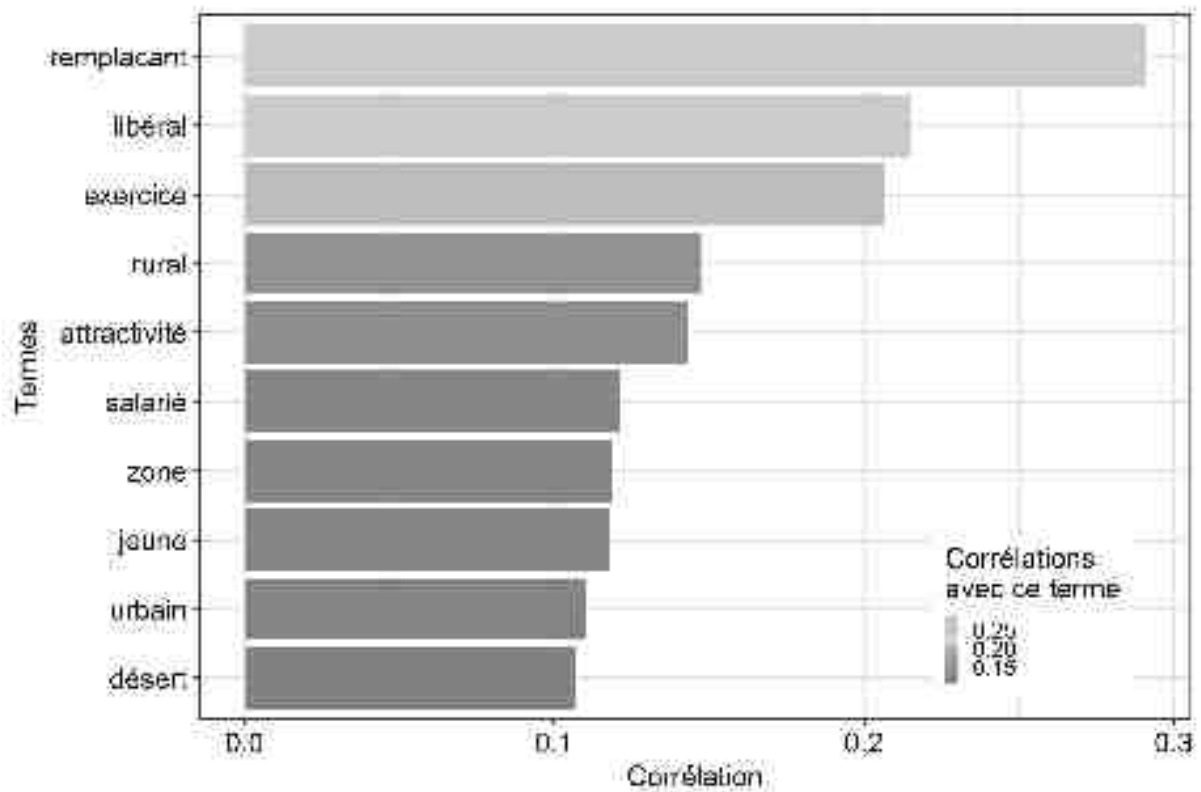
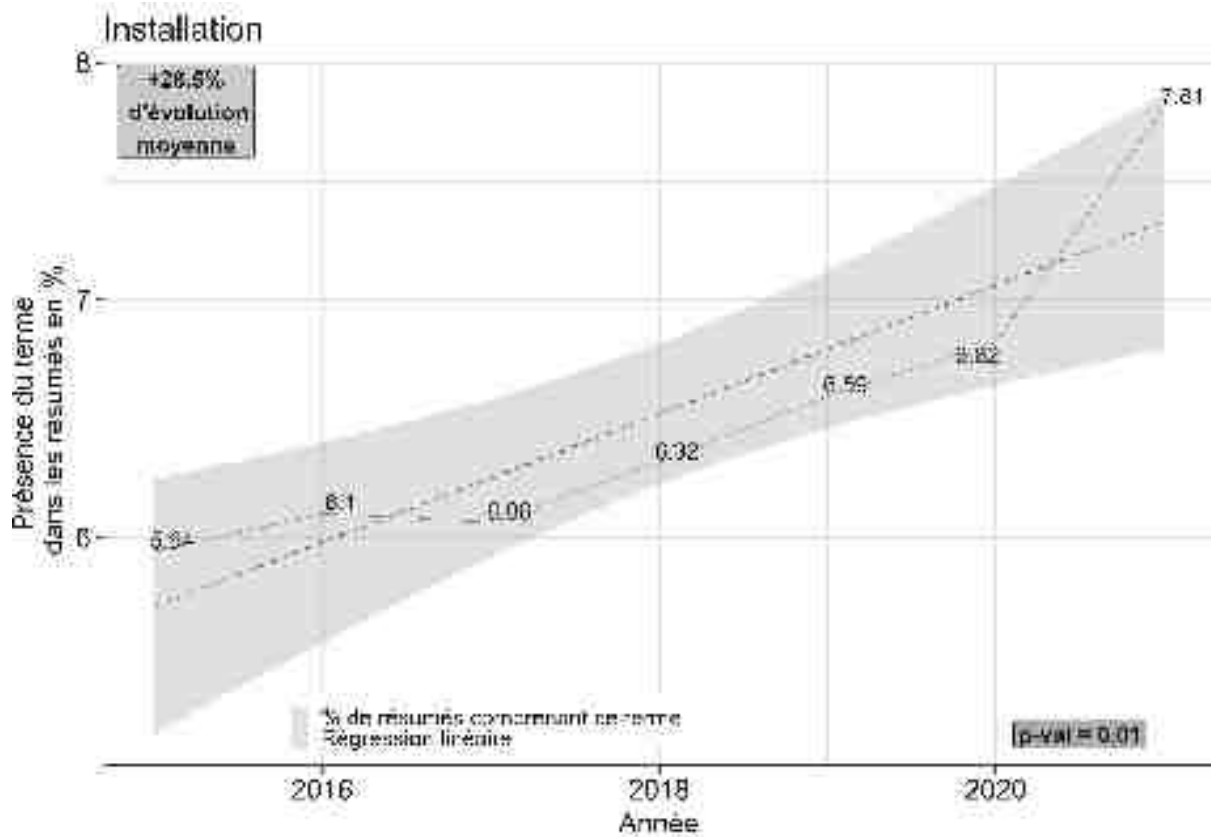


Figure 38 : Évolution et corrélations du terme installation sur la période 2015 à 2021

Sur la Figure 38, nous observons l'évolution du terme installation, que l'on a choisi pour représenter l'installation et l'exercice de la médecine générale en milieu rural. Nous pouvons y observer une augmentation de +28,5% ($p=0,01$).

La problématique de la démographie médicale peut transparaître au travers du graphique des corrélations avec les termes attractivité et désert. En effet cette problématique va s'accroître dans les prochaines années au vu de la baisse du nombre de médecins généralistes, qui était déjà de -10% de 2010 à 2020 (67).

3.4.1.13 La prévention en médecine générale

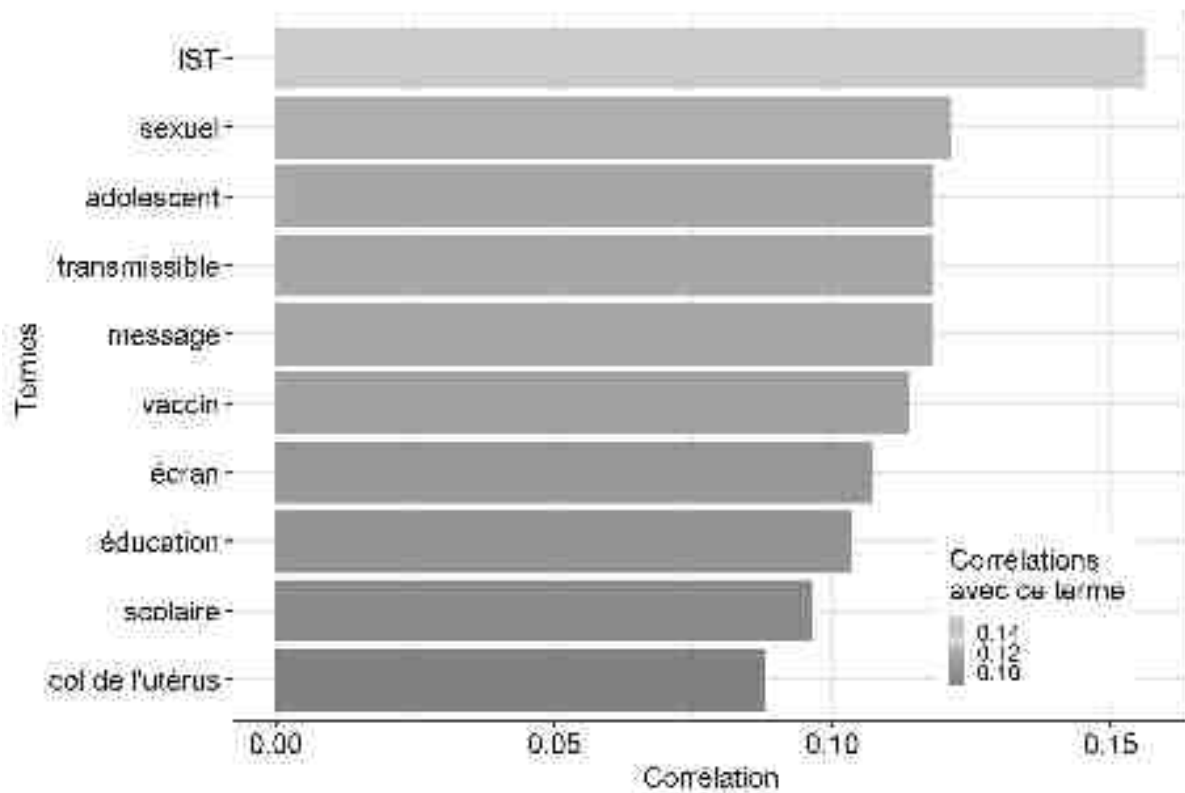
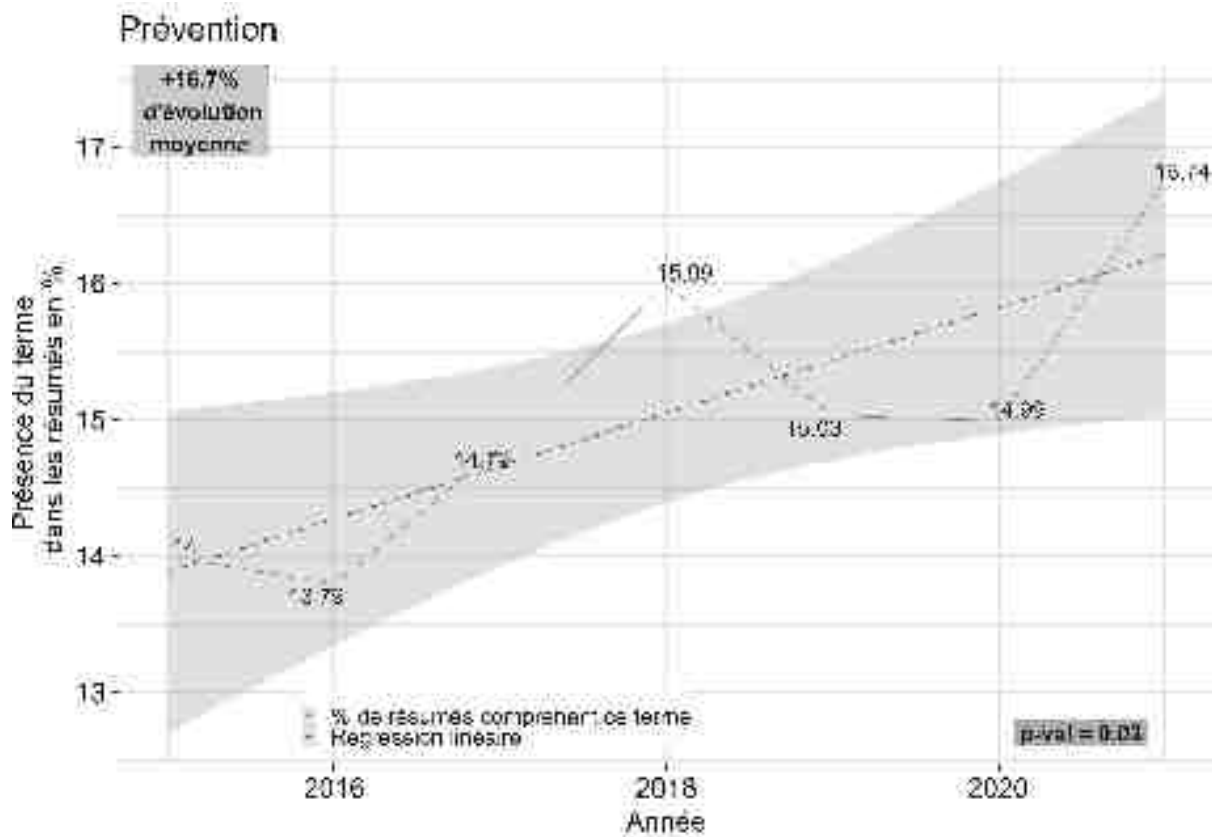


Figure 39 : Évolution et corrélations du terme prévention sur la période 2015 à 2021

Sur la Figure 39, nous observons l'évolution du terme prévention, que l'on a choisi pour représenter la prévention en médecine générale. Nous pouvons y observer une augmentation de +16,7% ($p=0,03$).

La prévention en médecine générale tient une place importante. Sur le graphique des corrélations nous pouvons observer des termes comme vaccin et écran faisant eux-mêmes partis d'un thème en forte augmentation (Figure 32 et Figure 27). Ces deux composantes sont donc en augmentation au sein de la médecine préventive. La thématique générale de la prévention semble donc croissante en termes d'intérêt porté par les internes de médecines générales et notamment les deux axes que sont les écrans et la vaccination.

3.4.2 Résultats avec diminution significative

3.4.2.1 Hospitalisation et personnes âgées

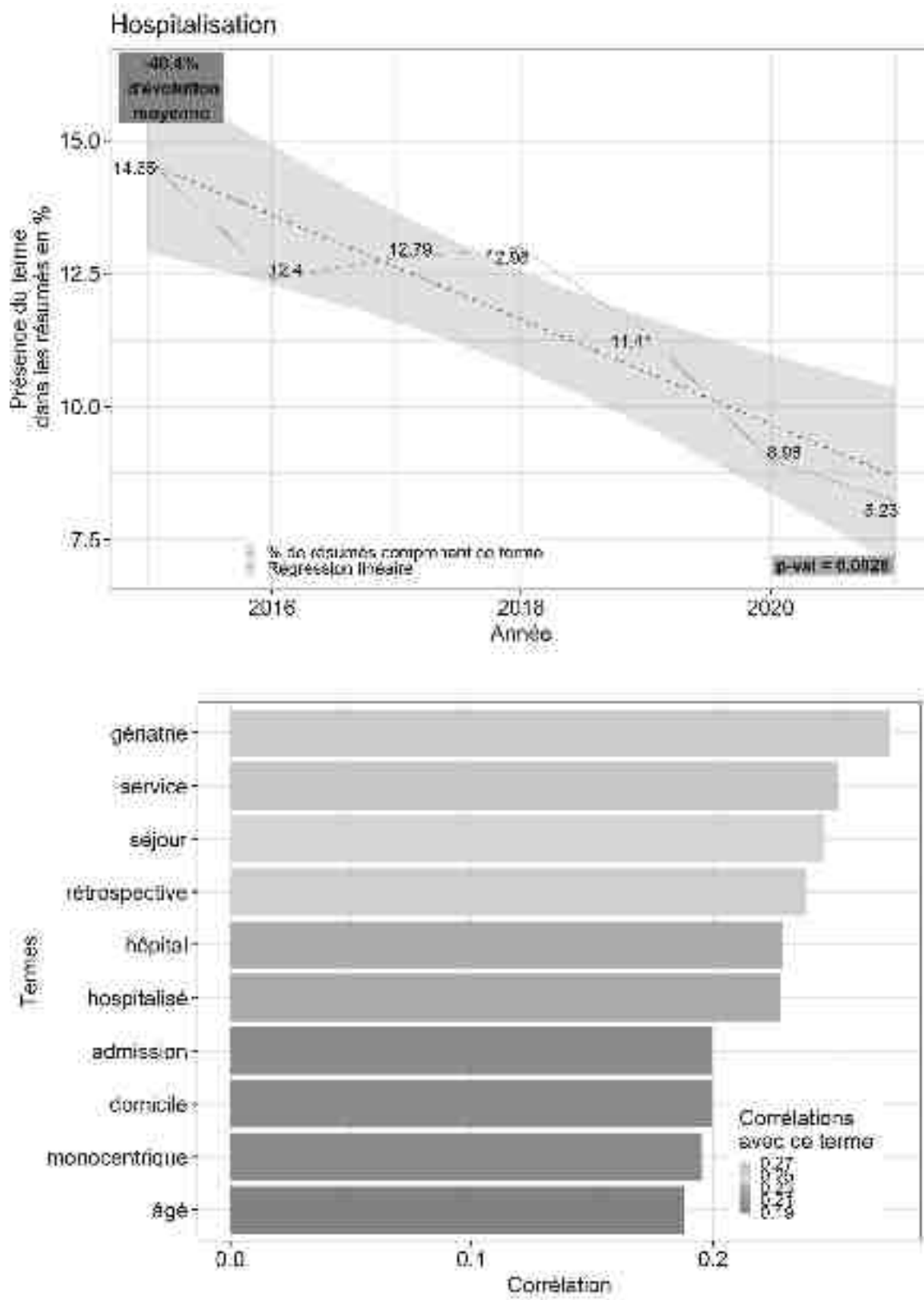


Figure 40 : Évolution et corrélations du terme hospitalisation sur la période 2015 à 2021

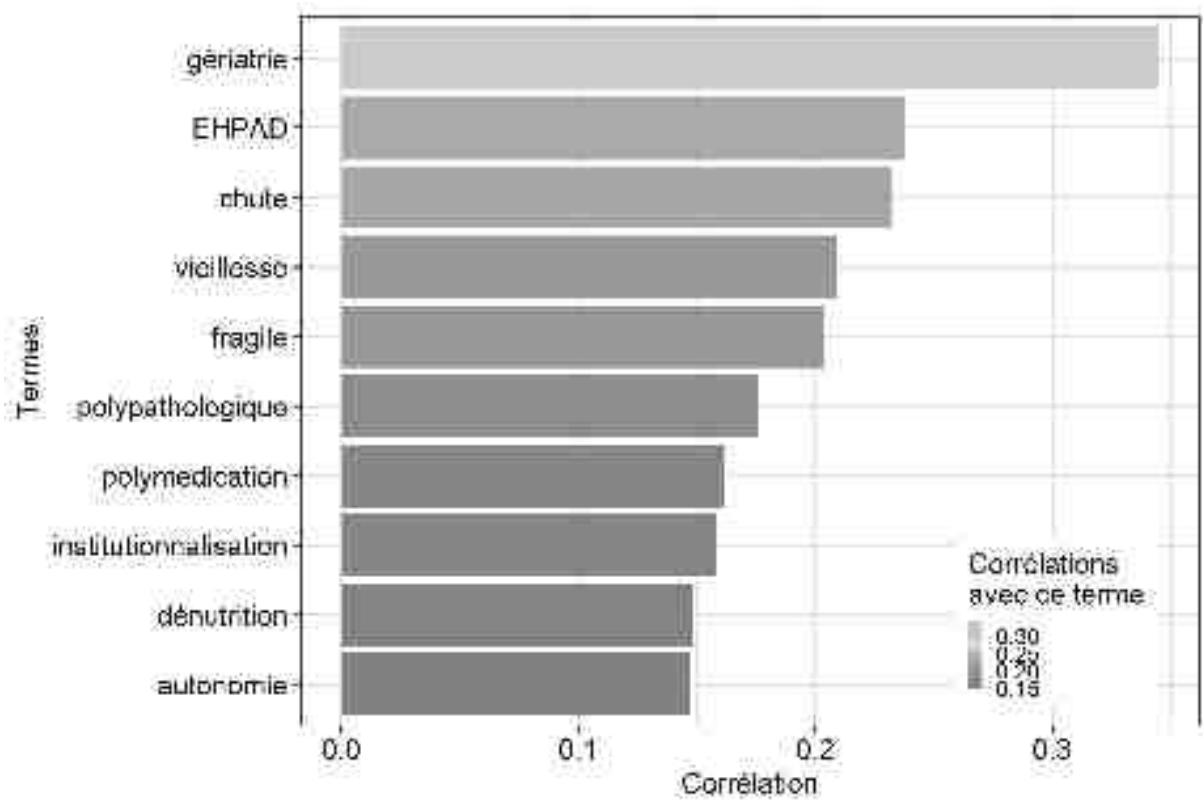
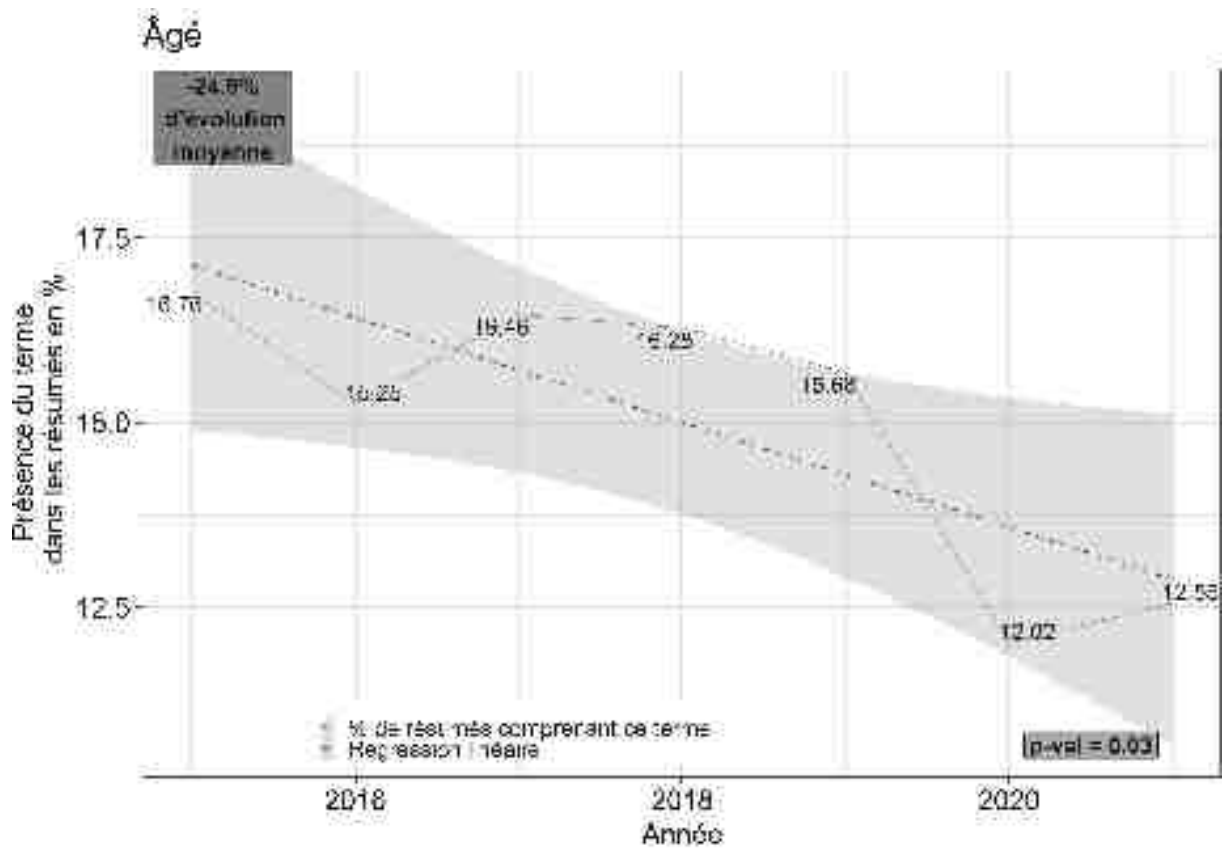


Figure 41 : Évolution et corrélations du terme âgé sur la période 2015 à 2021

Sur les Figure 40 et Figure 41, nous observons l'évolution des termes « hospitalisation » et « âgé », représentatifs des hospitalisations et des personnes âgées. Nous pouvons y observer une diminution marquée de respectivement -40,4% ($p=0,0028$) et de -24,9% ($p=0,03$).

Nous présentons ces deux figures ensemble, car les résultats sont intriqués comme nous pouvons l'observer sur le graphique des corrélations où le terme gériatrie apparaît en première position pour les deux termes. Ces résultats reflètent les changements de paradigmes des sujets de thèses de médecine générale. Les thèses sont plus centrées sur des problématiques de médecine générale et moins sur l'hospitalisation. Nous pouvons déjà observer cette tendance dans les années 1997 à 99 et 2007 à 2009 (43) dans cette thèse comparant les sujets de médecine générale à ces deux époques.

Le sujet âgé étant corrélé au contexte d'hospitalisation, l'hypothèse que l'on pourrait émettre est que comme les études à l'hôpital diminuent, il y a mécaniquement moins d'étude sur leur population principale (68). Alors que parallèlement la population de personnes âgées augmente : « *La part de personnes âgées de 65 ans ou plus a augmenté de près d'un quart en l'espace de 10 ans, passant de 17 % en 2012 à 21 % en 2022* » (69).

3.4.2.2 Soins à domicile

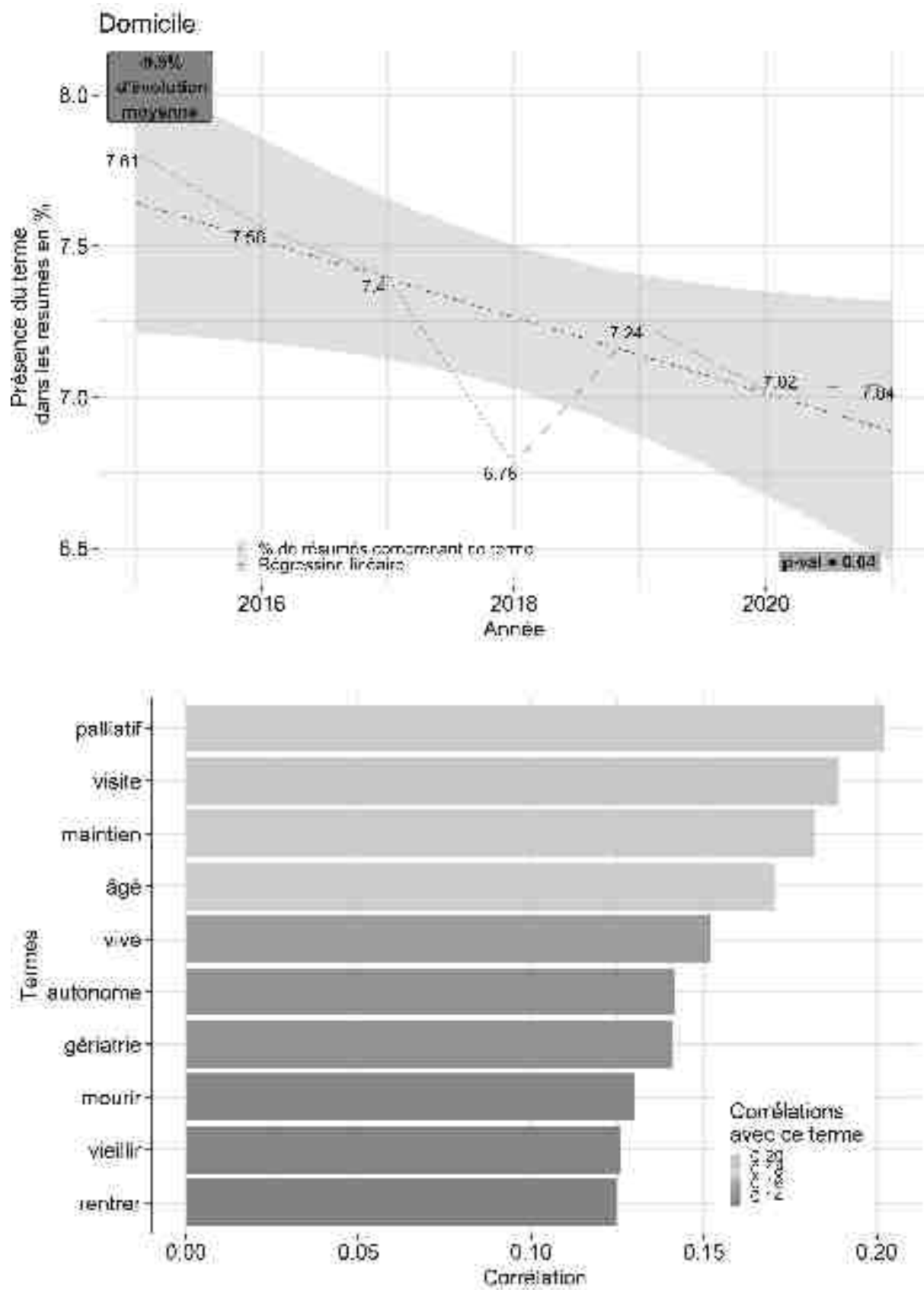


Figure 42 : Évolution et corrélations du terme domicile sur la période 2015 à 2021

Sur la Figure 42, nous observons l'évolution du terme domicile, que l'on a choisi pour représenter les soins à domicile en médecine générale. Nous pouvons y observer une diminution de 9,9 % ($p=0,04$).

Ici nous observons que le terme est en corrélation avec les personnes âgées, le maintien au domicile, les soins palliatifs, et le décès à domicile.

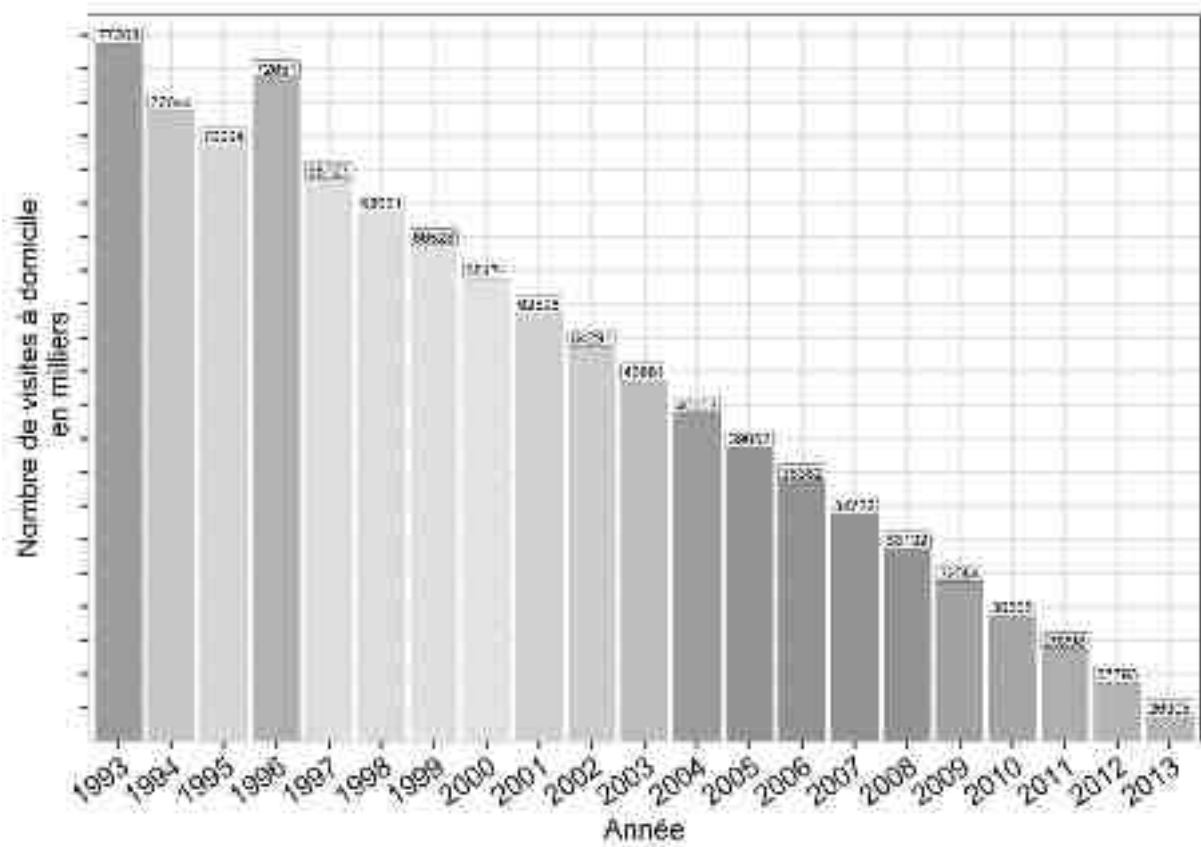


Figure 43 : Nombre de visites à domicile en France de 1993 à 2013, d'après les données d'éco-santé (70)

Figure 43, nous remarquons que cette diminution est déjà présente dès 1993, se poursuit jusqu'en 2013. En 2016 les chiffres étaient encore à la baisse avec 24 442 milliers de visites à domicile. Depuis les années 1980, on observe globalement une augmentation des consultations et une diminution des visites à domicile (71).

3.4.3 Résultats avec augmentation non significative

Nous ne détaillerons pas les résultats avec augmentation non significative ; cependant ils restent consultables en annexe, notamment pour le contexte qu'ils peuvent apporter à certaines évolutions observées.

Ci-dessous la liste de ces résultats en augmentation non significatives :

- Infections sexuellement transmissibles : + 109,5 % (Annexe 6, p150).
- Maltraitance : + 103,5 % (Annexe 7, p151).
- Anxiété : + 47,1 % (Annexe 8, p152)
- Dépression : + 11,9 % (Annexe 9, p153)
- Stress : + 52,3 % (Annexe 10, p154)
- Épuisement professionnel : + 39,5 % (Annexe 11, p155)
- Sommeil : + 35,6 % (Annexe 12, p156)
- Travail : + 18,5 % (Annexe 13, p157)
- Adolescent : + 25,5 % (Annexe 14, p158)
- Surpoids : + 16,2 % (Annexe 15, p159)
- Obésité : + 14,5 % (Annexe 16, p160)
- Femme : + 9,8 % (Annexe 17, p161)
- Formation : + 3,9 % (Annexe 18, p162)
- Pédiatrie : + 3,9 % (Annexe 19, p163)
- Tabagisme : + 1,3 % (Annexe 20, p164)

3.4.4 Résultats avec diminution non significative

Nous ne détaillerons pas les résultats en diminution non significative ; cependant ils restent consultables en annexe, notamment pour le contexte qu'ils peuvent apporter à certaines évolutions observées.

Ci-dessous la liste de ces résultats en diminution non significatives :

- Urgence : - 21,7 % (Annexe 21, p165)
- Gériatrie : - 35,4 % (Annexe 22, p166)
- Diabète : - 22,3 % (Annexe 23, p167)
- Prescription : - 14 % (Annexe 24, p168)
- Palliatif : - 12,8 % (Annexe 25, p169)
- Homme : - 4,1 % (Annexe 26, p170)
- Sevrage : - 3 % (Annexe 27, p171)
- Alcool : - 5,1 % (Annexe 28, p172)
- Grossesse : - 3,7 % (Annexe 29, p173)
- Douleur : - 2 % (Annexe 30, p174)

3.5 Résultats par comparaison directe en LDA

Tableau 4: Résultats de la LDA sur la période 2015 à 2021, avec l'évolution moyenne par année (* $p < 0,05$)

Les 8 premiers termes racinisés classés par allocation latente de Dirichlet	Interprétation	2015	2021	Évolution moyenne par année
vih.covid.réunion.épidem.risqu.infect.cris.sanitaire	Les risques pandémiques	1,7 %	2,9 %	1,103x
entretien.médecin.général.sem.qualit.dirig.aupres.individuel	<u>Étude qualitative</u>	<u>4,9 %</u>	<u>6,9 %</u>	<u>1,071x*</u>
utilis.outil.médecin.médical.commun.inform.amélior.téléconsult	<u>Outil informatique</u>	<u>2,1 %</u>	<u>3%</u>	<u>1,07x*</u>
vaccin.couvertur.contr.hpv.inform.anti.gripp.antigrippal	<u>Vaccination</u>	<u>1,3 %</u>	<u>1,9 %</u>	<u>1,064 x*</u>
troubl.psychiatr.dépress.sommeil.stress.mental.professionnel.qualit	<u>Santé mentale</u>	<u>1,6 %</u>	<u>2,5 %</u>	<u>1,063x*</u>
revu.littératur.articl.recherch.don.bas.systémat.analys	<u>Revue de la littérature</u>	<u>2%</u>	<u>2,8 %</u>	<u>1,059x*</u>
social.soin.sant.précar.acces.situat.popul.person	<u>Social</u>	<u>1,8 %</u>	<u>2,4 %</u>	<u>1,059x*</u>
sant.soin.professionnel.commun.infirmi.coordin.projet.collabor	<u>Réseaux de soins</u>	<u>2,7 %</u>	<u>3,9 %</u>	<u>1,059x*</u>
violenc.victim.femm.repérage.médecin.conjugal.maltrait.dépistag	<u>Violences</u>	<u>1,3 %</u>	<u>1,7 %</u>	<u>1,059x*</u>
physiqu.activ.obes.poid.alimentair.surpoid.sant.sport	<u>Surpoids et de l'obésité</u>	<u>1,7 %</u>	<u>2,1 %</u>	<u>1,05x*</u>
outil.évalu.utilis.sit.fich.valid.expert.item	<u>Méthodologie</u>	<u>2,1 %</u>	<u>3%</u>	<u>1,046x*</u>
enfant.parental.adolescent.pédiatr.âge.consult.nourrisson.écran	Suivi pédiatrique	1,6 %	2,1 %	1,031x
médecin.relat.soign.vécu.représent.médical.patient.malad	Relation médecin-patient	3,8 %	4,1 %	1,027x
femm.contracept.sexual.gynécolog.ivg.homm.ist.diu	Gynécologie et contraception	2%	2,3 %	1,025x
inform.connaiss.prévent.sant.questionnair.sujet.avoir.internet	Médecine préventive	3%	3,3 %	1,024x
tabag.consomm.alcool.sevrage.addictolog.arrêt.usag.cigaret	Sevrage alcool et tabac	1,4 %	1,6 %	1,015x
médecin.général.questionnair.pratiqu.format.recommand.connaiss.répondr	Formation des MG	4,7 %	5,1 %	1,007x
médecin.patient.consult.général.trait.questionnair.motif.cabinet	Pratiques des MG	3,1 %	3,2 %	1,006x
appel.médical.soin.régul.urgenc.demand.gard.samu	Régulation des soins non programmées	1,7 %	2%	1,003x
femm.grossess.allait.mer.enceint.maternel.matern.suiv	Grossesse et maternité	1,4 %	1,7 %	1,002x
charg.pris.médecin.général.soin.trait.patient.difficult	Prises en charges	3,5 %	3,4 %	0,996x
format.pratiqu.particip.compétent.techniqu.form.gest.connaiss	Formation technique	2,1 %	2%	0,992x
facteur.âge.scor.associ.signif.analys.entre.statist	Facteurs de risques	3,2 %	3,1 %	0,988x
aid.soin.domicil.palliat.patient.fin.anticip.direct	Soins à domicile et fin de vie	2,1 %	1,9 %	0,986x
sport.militair.traumat.pratiqu.lésion.fractur.blessur.accident	Traumatologie	1,8 %	1,5 %	0,984x
intern.stag.médecin.général.étudi.enseign.format.univers	Formation des internes	2,3 %	2,3 %	0,978x
médecin.général.install.exercic.libéral.médical.cabinet.remplac	Installation et exercice de la MG	3,3 %	3,2 %	0,975x
dépistag.canc.test.âge.femm.particip.réalis.sein	Dépistage cancer	1,7 %	1,7 %	0,973x
douleur.patient.trait.utilis.efficac.antalg.chroniqu.effet	Douleurs chroniques	2%	1,6 %	0,971x
group.signif.différent.compar.mois.criter.apres.deux	Méthodologie	2,9 %	2,4 %	0,969x
the.and.method.resultat.introduct.materiel.for.with	Méthodologie	0,9 %	0,6 %	0,968x
diagnostic.examen.échograph.cliniqu.réalis.ecg.biolog.test	Outils diagnostiques	2,3 %	1,8 %	0,968x
patient.diabet.thérapeut.éduc.malad.trait.etp.observ	<u>Diabète</u>	<u>2,3 %</u>	<u>1,8 %</u>	<u>0,957x*</u>
patient.cardiaqu.mortal.insuffis.respiratoir.rénal.réanim.bpc	Pathologies chroniques	2%	1,4 %	0,952x
antibiothérap.infect.prescript.recommand.urinair.résist.cas.bactérien	Antibiothérapies	2%	1,2 %	0,948x
prescript.médic.trait.effet.pharmacien.ordon.indésir.indiqu	<u>Prescriptions</u>	<u>2,7 %</u>	<u>1,7 %</u>	<u>0,943x*</u>
diagnostic.cas.patholog.cliniqu.symptôm.malad.âge.fréquent	<u>Facteurs de risque</u>	<u>2,7 %</u>	<u>2%</u>	<u>0,943x*</u>
patient.hospitalis.mois.suiv.chirurg.jour.sort.dur	<u>Hospitalisation</u>	<u>2,8 %</u>	<u>1,8 %</u>	<u>0,939x*</u>
patient.cardiovasculair.risqu.vasculair.avc.artériel.chez.hta	<u>FDR cardiovasculaire</u>	<u>2%</u>	<u>1,3 %</u>	<u>0,932x*</u>
urgenc.patient.charg.servic.pris.hospitali.admiss.hospitalis	<u>Urgences et hospitalisations</u>	<u>3,8 %</u>	<u>2,1 %</u>	<u>0,929x*</u>
âgé.person.gériatr.âge.ehpad.fragil.chut.sujet	<u>Personne âgée fragile</u>	<u>2,9 %</u>	<u>1,8 %</u>	<u>0,925x*</u>
trait.anticoagul.patient.hémorrag.sous.risqu.aod.vitamin	<u>Anticoagulants</u>	<u>2,4 %</u>	<u>0,9 %</u>	<u>0,859x*</u>

Sur le Tableau 4, nous observons les résultats obtenus après LDA sur l'ensemble du corpus et la modélisation de l'évolution moyenne annuelle par régression binomiale.

Les thèmes non méthodologiques qui sont prépondérants, c'est-à-dire ceux qui sont présents dans au moins 3 % des thèses sont :

1. La formation des médecins généralistes (4,7% en 2015 et 5,1% en 2021)
2. La relation médecin-patient (3,8% en 2015 et 4,1% en 2021)
3. La médecine préventive (3% en 2015 et 3,3% en 2021)
4. Les prises en charges (3,5% en 2015 et 3,4% en 2021)
5. Les pratiques des médecins généralistes (3,1% en 2015 et 3,2% en 2021)
6. L'installation et l'exercice de la médecine générale (3,3% en 2015 et 3,2% en 2021)
7. L'outil informatique (2,1% en 2015 et 3% en 2021)

Nous pouvons observer des augmentations significatives pour les études qualitatives et les revues de la littérature. Par ailleurs nous observons que pour les thématiques du surpoids et de l'obésité ainsi que de la santé mentale, nous mettons en évidence une évolution positive significative. A noter le thème du risque pandémique qui perd en significativité par l'apparition et l'augmentation du sous thème du covid sur les années 2020 et 2021. Le reste des thématiques en augmentation significative reste semblable à ce que l'on a pu observer précédemment.

En ce qui concerne les thématiques en diminution nous pouvons principalement observer une nouvelle thématique non précédemment décrite qui est celle des traitements anticoagulants qui est en diminution forte depuis 2015. Les thématiques du diabète, des prescriptions et des urgences sont en diminution significative contrairement à précédemment où la diminution était observée de façon non significative. Le reste des thématiques en diminution significative reste semblable à ce que l'on a pu observer précédemment.

4 Discussion.

Il est possible de déterminer les thèmes des résumés de thèses avec des outils de fouilles de textes et de quantifier leur proportion. Pour ce qui est de l'évolution de ces sujets de thèse, il est possible dans certains cas de déterminer de façon significative s'ils sont en augmentation ou en diminution.

4.1 Discussion des principaux résultats

Pour cette partie nous nous aiderons des programmes de cours du Diplôme de formation approfondie en sciences médicales de l'université de Aix-Marseille(72) ainsi que du programme de cours de l'université de Poitiers concernant le DES de médecine générale(73). Ce choix étant simplement motivé par leur disponibilité et la mention de la quantité d'heures de cours par thèmes, des études spécifiques par facultés seraient nécessaires pour évaluer de façon plus formel les adéquations entre les tendances de ces sujets et leurs enseignements en deuxième et troisième cycle.

4.1.1 Téléconsultation

Plusieurs facteurs ont pu contribuer à l'essor de l'utilisation du terme « téléconsultation » dans résumés de thèses de médecine générale de 2016 à 2021.

Les téléconsultations nécessitent un accès Internet de qualité et des équipements techniques performants tels que des ordinateurs, des tablettes et des smartphones. Ces technologies se sont considérablement améliorées et sont devenues plus répandues au cours de cette période(74,75), cela pourrait expliquer l'intérêt croissant pour la téléconsultation.

Des réformes ont été faites pour faciliter le développement de la téléconsultation, ce qui peut expliquer l'intérêt croissant pour ce traitement. A partir de 2016 des expérimentations sur le territoire français ont été favorisées par la création cotation d'actes spécifiques remboursables par la sécurité sociale et généralisée sur tout le territoire en 2018(76).

L'émergence de la pandémie de COVID-19 au cours de cette période, amenant les gens à chercher des alternatives aux consultations en face à face, a elle aussi favorisé le développement des téléconsultations(77).

La téléconsultation peut être perçue comme une modalité de soins pratique, efficace et accessible, ce qui peut expliquer l'intérêt accru pour cette approche. En effet, la crise de la démographie médicale actuelle et les difficultés croissantes pour obtenir un rendez-vous(78) peuvent favoriser son développement.

Au niveau des enseignements pouvant traiter de ce sujet en deuxième cycle (items "Télémédecine, télésanté et téléservices en santé" et "Santé et numérique") ils représentent 2 heures de cours, soit 0,3% de la durée totale des cours du DFASM à Aix-Marseille. Nous n'avons pas retrouvé de cours obligatoires sur ce sujet en TCEM à Poitiers.

4.1.2 Écrans

L'augmentation de ce terme dans les thèses semble corrélée au sujet de la téléconsultation et partage en partie les mêmes arguments pour expliquer son augmentation. Grâce aux progrès technologiques et à la démocratisation de l'accès à Internet, les écrans deviennent de plus en plus performants et accessibles(74,75). Aussi la crise du Covid a favorisé les usages des écrans via les téléconsultations mais aussi le télétravail.

Hormis ces arguments il semble que l'on s'intéresse davantage aux effets des écrans chez les enfants. En effet avec un accès de plus en plus facilité à ces technologies, les enfants peuvent être exposés plus facilement à des effets potentiellement négatifs. Premièrement, il existe un risque d'addiction aux écrans qui peut entraîner des troubles du sommeil, de l'attention et de l'humeur(2). L'utilisation excessive d'écrans peut également affecter le développement de la motricité fine et de la coordination œil-main, ainsi que les compétences de socialisation et de communication verbale. Cela peut augmenter le risque d'obésité et de troubles visuelles. Enfin, il existe un risque d'exposition excessive à des contenus violents, sexuels ou inappropriés, qui peuvent nuire au développement

émotionnel et moral d'un enfant. Cependant, il est important de noter que l'utilisation des écrans peut également profiter aux enfants, à condition qu'elle soit modérée et supervisée.

Les écrans peuvent être utilisés pour apprendre de nouvelles choses et développer de nouvelles compétences. Les enfants peuvent utiliser des jeux éducatifs pour apprendre de nouvelles langues ou résoudre des problèmes de logiques. Grâce aux écrans, les enfants peuvent rester en contact avec leurs amis et leur famille même lorsqu'ils sont éloignés. Ceci est particulièrement utile pour les enfants qui vivent dans des régions éloignées, qui voyagent beaucoup ou bien sûr en cas d'urgence. Les écrans permettent aussi aux enfants de pouvoir se divertir de façon plus autonome (2,75,79,80).

Au niveau des enseignements pouvant traiter de ce sujet en deuxième cycle nous pouvons trouver « Prévention de l'exposition aux écrans ». Il fait partie d'un groupe de cours qui représentent 3 heures de cours, soit 0,5% de la durée totale des cours du DFASM à Aix-Marseille. Nous n'avons pas retrouvé de cours obligatoires sur ce sujet en TCEM à Poitiers.

4.1.3 COVID

La pandémie de COVID-19 a eu un impact considérable sur la santé et le bien-être. La présence croissante du terme dans la littérature reflète l'importance de la pandémie et l'intérêt pour l'étude de différents aspects de la COVID-19, notamment l'impact sur les soins de santé généraux(56,77,81).

Il n'a pas été retrouvé d'enseignement spécifique de cette pathologie mais il est peu probable que les facultés ou les CHU n'ont pas mise en place des formations spécifiques durant cette période.

4.1.4 Violence

Nous constatons une augmentation de l'intérêt pour le thème des violences dans les thèses de médecine générale en France de 2015 à 2021. Cette évolution est cohérente avec les chiffres du gouvernement qui déclare que : « le nombre de victimes de coups et blessures volontaires (sur personnes de 15 ans ou plus) enregistrées augmente très fortement en 2021 (+12 %, après +1 % en

2020 et +8 % en 2019) » (57). La violence conjugale reste au premier plan comme nous le constatons sur le graphique des corrélations, mais nous pouvons observer que les faits de violences ne concernent pas seulement la sphère conjugale en médecine. En dernière position se retrouve la maltraitance, qui est une forme de mauvais traitement envers des personnes vulnérables. Nous pourrions constater l'augmentation non significative de cette dernière (Annexe 7 : Évolution et corrélations du terme maltraitance sur la période 2015 à 2021, page 140).

Il est possible que l'incidence de la violence ait réellement augmenté au cours de cette période ou alors qu'elle soit de plus en plus signalée. De plus, de nouvelles politiques ou programmes de lutte contre les violences et le sexisme ont été mis en place durant cette période(19,82–84). La pandémie de COVID et le confinement résultant pourrait aussi avoir été un facteur d'augmentation de violences conjugales(85).

Les huit heures de cours consacrées aux violences (« *Violences et santé* », « *Violences sexuelles* », « *Violences et maltraitances aux adultes* », « *Maltraitance et enfants en danger* », « *Dépistage et prévention des violences faites aux femmes* », « *Suspicion de maltraitance et enfant en danger* ») représentent 1,35% de la durée totale de cours du DFASM à Marseille.

A cela s'ajoute « Le patient vivant dans un contexte de violence : entourage, profession », un cours de 2 heures, en deuxième année de TCEM.

On constate ici que la formation sur ce sujet a été récemment renforcée, ce qui est cohérent avec les résultats que nous avons pu observer et la situation actuelle en France sur les violences.

4.1.5 Parcours

L'augmentation du terme "parcours" dans les résumés de thèses de médecine générale en France de 2015 à 2021, ici faisant principalement référence au parcours de soin peut être dû à la promotion de celui-ci. Le parcours de soins coordonnés vise à améliorer la qualité et la continuité des soins en encourageant les patients à consulter leur médecin traitant en premier lieu et en favorisant

une meilleure coordination entre différents professionnels de santé impliqués dans leur prise en charge. La loi du 13 août 2004 qui a institué le parcours de soins coordonnés, donne au médecin traitant un rôle de coordonnateur de soin, avec pour incitation un meilleur remboursement pour les patients respectant ce parcours(58). D'autres arguments peuvent aussi être avancé comme par exemple la complexité croissante du système de santé(86), la nécessité d'efficience devant la crise de la démographie médicale(87) ainsi que l'essor des nouvelles technologies de communication(74) qui peuvent faciliter l'adoption et le développement du parcours de soin.

Ce thème étant très transversal nous n'avons pas pu identifier clairement des enseignements le traitant en DFASM, cependant un séminaire de 3 heures y est consacré en TCEM à Poitier (« Continuité / coordination des soins, réseaux »).

4.1.6 Social

L'augmentation du terme "social" dans les résumés de thèses de médecine générale en France de 2015 à 2021 pourrait être la conséquence d'une nouvelle prise de conscience de l'importance d'une prise en soins globale et respectant les attentes du patient. Les facteurs sociaux tels que le niveau d'éducation, le niveau de revenu, le logement, etc., ont un impact sur la santé des individus. Il est important de prendre en compte l'ensemble des besoins des patients, y compris les besoins sociaux(61,88).

Les 3 heures de cours consacrées à la thématique sociale (« Prescription médicale chez un patient en situation de précarité », « Situation sociale précaire et isolement) représentent 0,5% de la durée totale de cours du DFASM à Marseille.

4.1.7 Vaccination

Nous avons pu observer que le terme vaccination augmentait de 2015 à 2021 dans les résumés de thèses de médecine générale. Cette augmentation peut paraître paradoxale car la vaccination est une des inventions de la médecine la plus efficace en termes de prévention des décès et une des plus ancienne.

En effet l'invention de la vaccination remonte à 1796 par Edward Jenner contre la variole, à noter que des techniques similaires à celle-ci peuvent être retrouvées dans des textes de médecine chinoise sans pour autant que ces techniques soient largement développées(89). Depuis, l'invention de nouveaux vaccins s'est accélérée de plus en plus, avec en moyenne un nouveau vaccin tous les 20 ans au XIXème siècle, tous les 4 ans et demi au XXème siècle et au début du XXIème siècle à nos jours un nouveau vaccin au maximum tous les deux ans et demi(89–91).

Donc une des hypothèses que l'on peut émettre est que comme il y a de plus en plus de nouveaux vaccins, il y a plus de publications de thèses sur ce thème.

Selon l'OMS, la vaccination permet d'éviter 3,5 à 5 millions de décès annuellement dus à des maladies telles que la diphtérie, le tétanos, la coqueluche, la grippe et la rougeole(92).

Cependant la couverture vaccinale en France est loin d'être optimale comme par exemple contre la grippe, le pneumocoque, la coqueluche ou le papillomavirus(93–95). Il y a donc des améliorations possibles en matière de couverture vaccinale et c'est probablement une des raisons pour laquelle ce thème est de plus en plus traité dans les thèses de médecine générale.

Un des obstacles possibles à l'obtention d'une couverture vaccinale correcte est bien sûr l'adhésion du patient à la vaccination. Or depuis quelques années on observe une défiance latente vis-à-vis des vaccins, compliquant les prises en soin chez certains patients. Une des explications possible est la conséquence de publications mettant en doute à tort le rapport bénéfice risque de la vaccination (96,97) sur principalement les vaccins rougeole-oreillons-rubéole et hépatite B. De là peut découler un biais cognitif, qui est un raisonnement distordu pouvant conduire à des idées fausses sur les risques et les avantages de la vaccination. Elle peut également être influencée par des facteurs sociaux tels qu'une crise de confiance envers les autorités sanitaires et la diffusion rapide de fausses informations ou de fausses informations sur les vaccins sur les réseaux sociaux et les médias(62).

Il faut donc pouvoir trouver des approches pour faire adhérer ces patients à la vaccination, ce qui peut faire l'objet de thèses.

Certaines mesures de santé publique sur l'extension des recommandations vaccinales peuvent aussi stimuler la production de thèses sur ce thème, par exemple l'extension de la vaccination contre HPV aux garçons.

De façon assez paradoxale nous n'avons pas retrouvé d'heures de cours au programme du DFASM de l'université Aix Marseille et pas non plus sur le programme de TCEM à Poitiers. Cet item est cependant au programme et traité de façon complète dans le livre de référence Pilly édition 2020, incluant aussi une partie sur l'hésitation vaccinale.

4.1.8 Sport

L'augmentation de ce thème peut être réactionnelle à certaines mesures de santé publique comme par exemple la mise en place du sport sur ordonnance(98). Au total, les heures de cours ayant pour relation la pratique de l'exercice physique ont été estimées à 5 heures de cours ce qui représente environ 0,8% de la durée totale des cours du DFAS à Marseille.

4.1.9 Allaitement

L'augmentation de 47,7% du terme "allaitement" dans les résumés de thèses de médecine générale en France de 2015 à 2021 peut être dû à une prise de conscience de l'importance de la promotion et du conseil à l'allaitement. L'allaitement peut avoir des bénéfices pour la santé de la mère et de l'enfant des études ont démontré que l'allaitement peut réduire le risque de développer certaines maladies chroniques. De plus, il est associé à un meilleur développement cognitif des enfants. Il est recommandé de pratiquer un allaitement maternel exclusif et prolongé pendant au moins 4 à 6 mois pour bénéficier de ces effets positifs. Cependant, le taux d'allaitement maternel en France est relativement faible et sa durée moyenne est de 15 semaines. La formation des professionnels de santé et la sensibilisation des conjoints, peuvent contribuer à augmenter le taux d'initiation et la durée de l'allaitement maternel. Il est important que les professionnels de santé soient en mesure de fournir une aide et des conseils adaptés pour aider les mères à réussir leur allaitement. Cela peut inclure le soutien pour surmonter les difficultés courantes comme les douleurs et les lésions

du mamelon, l'engorgement mammaire et la mastite, ainsi que l'aide à la prise de décision en cas de problèmes de santé chez la mère ou l'enfant(1,4,65,99,100).

Une heure de cours son consacré à l'allaitement normal et difficile, ce qui représente environ 0,17% des 594 heures totales de cours en DFASM.

Il n'y a pas de correspondance sur le programme de cours de TCEM de Poitier en 2018.

4.1.10 Parentalité et enfant

Cette augmentation des concepts de parentalité et d'enfant en médecine générale peut être en partie attribuée aux réformes de 2017 qui ont élargi les stages en gynécologie et en pédiatrie et ont offert davantage de possibilités de stages en médecine générale(101). Cela peut signifier que les étudiants en médecine s'intéresseront davantage à ces domaines et mettront davantage l'accent les soins aux enfants, la santé de la femme et le rôle parental dans leur formation et leur pratique future.

Nous n'avons pas réussi à estimer la part de ces thèmes dans le cursus de DFASM et TCEM au vu de leur transversalité.

4.1.11 Contraception

Près de 80% des femmes de 15 à 49 ans utilisent une méthode contraceptive(66), ce qui peut expliquer l'engouement de ce thème en médecine générale. Il est important de souligner que la contraception n'est pas seulement une affaire de femmes. Les hommes ont également un rôle à jouer dans la prévention des grossesses non désirées en utilisant des méthodes de contraception comme les préservatifs. Comme précédemment, nous pouvons y voir une conséquence de la réforme de 2017 qui a élargi les stages en gynécologie et en pédiatrie et a offert davantage de possibilités de stages en médecine générale(101).

Les modalités de prises en charge et d'accès à la contraception ont aussi été changé avec récemment la possibilité de délivrer gratuitement des préservatifs pour les moins de 26 ans (102), le remboursement de certains dispositifs de contraception féminin et des soins liés pour les moins de 26

ans depuis 2022 (103) qui était auparavant appliquée aux mineures (filles de 15 à 18 ans en 2013 et en janvier 2020, pour les filles de moins de 15 ans). Aussi, depuis le 1er janvier 2023, la contraception d'urgence hormonale peut être obtenue gratuitement sans ordonnance en pharmacie de ville.

Le cours « Prescrire une contraception et contraception d'urgence » représente 0,17% de l'ensemble des heures de cours du DFASM que nous avons pu identifier. Il n'y a pas de formation spécifique en TCEM sur la contraception identifiable.

4.1.12 Prévention

De nombreux thèmes déjà commentés font partie du domaine de la prévention, c'est pourquoi nous pouvons y observer une augmentation, cependant cela traduit aussi une volonté de renforcer les actions de médecine préventive. Dans les dernières années, il a été mis en place des actions de prévention ou de dépistage organisé ou non, nous pouvons citer par exemple :

- Des programmes de dépistage et de prévention de certains types de cancer (par exemple, dépistage du cancer du sein ,du colon, ou du col de l'utérus)(104)
- Des campagnes de sensibilisation sur l'importance d'une alimentation équilibrée, de l'activité physique, et de la lutte contre l'obésité(105)

Mais aussi de nouvelles campagnes de vaccination, de prévention des violences, des infections sexuellement transmissibles, du suicide et de la dépression, du tabagisme et du tabagisme passif ainsi que de nombreuses études(94,105–109).

Le nombre d'heures de cours de prévention représente environ 1,68% (10h) des heures de cours du DFASM.

4.1.13 Installation

L'augmentation de l'intérêt sur le thème de l'installation peut traduire le fait qu'en France, l'implantation de la médecine générale en milieu rural se heurte à de nombreux défis, comme la désertification médicale et le manque de médecins généralistes dans certaines régions. Il y a

actuellement une pénurie de médecins généralistes en France, en particulier en milieu rural, où il est souvent plus difficile d'attirer et de retenir des médecins(87). Il y a eu plusieurs initiatives visant à inciter les médecins à s'installer en milieu rural, comme les contrats d'engagement de service public (CESP). Le CESP créé le 21 juillet 2009 permet le versement d'une allocation mensuelle de 1 200€, les bénéficiaires s'engagent pendant un nombre d'années égal à celui durant lequel ils auront perçu l'allocation à exercer leurs fonctions dans une zone caractérisée par une offre de soins insuffisante(110). La loi de modernisation de notre système de santé du 26 janvier 2016 a prévu la création de zones de santé publique (ZSP) dans lesquelles les médecins généralistes bénéficient d'incitations financières(111).A cela peut s'ajouter différentes aides en fonction de la situation du praticien et du niveau de désertification médicale(112). La réforme de 2017 devrait consacrer une plus grande partie des stages en exercice extra hospitalier(101), cependant la formation à l'installation et à la gestion d'un cabinet en médecine libéral ne représente que 6 h en TCEM.

4.1.14 Santé mentale

Dans les résumés de thèses de médecine générale de 2015 à 2021 la représentation du thème santé mentale est en augmentation passant de 1,6% à 2,5%. Des améliorations sont en cours avec pour objectifs la feuille de route rédigé en 2018 qui vise à promouvoir le bien être mental, sur 3 axes principaux :

- Prévenir et repérer précocement la souffrance psychique, et prévenir le suicide.
- Garantir des parcours de soins coordonnés et soutenus par une offre en psychiatrie accessible, diversifiée et de qualité.
- Améliorer les conditions de vie et d'inclusion sociale et la citoyenneté des personnes en situation de handicap psychique(113).

Le développement de dispositifs de soutien psychologique à distance, tels que les plateformes de téléconsultation ou les lignes d'écoute téléphonique ont eux aussi pris de l'essor suite à la pandémie de covid 19(56,77).

La part de la santé mentale dans le cursus de médecine en DFASM et TCEM est difficilement évaluable.

4.1.15 Hospitalisation

Le thème « hospitalisation » est en décroissance de 2015 à 2021, observation que l'on peut aussi faire au niveau de l'analyse par LDA. Cette observation peut découler des modifications de la réforme de 2017 sur les prérequis pour les sujets de thèse de DES de médecine générale. En effet le sujet de la thèse doit être en relation directe avec la médecine générale. Or si l'on observe les résultats on peut constater que la baisse de ce thème intervient déjà avant la réforme, donc les étudiants « anciens régimes » n'ont pas pu être influencés par cela.

D'une manière générale, en prolongation de ce que l'on a pu observer dans les précédents résultats, on note une tendance au développement de la télémédecine, du parcours de soins coordonné et de la prévention. Ceci peut en réaction favoriser les prises en charges ambulatoires ou en journées partielles et donc diminuer la part des hospitalisations complètes(114). Au niveau pédagogique, en TCEM à Poitier, de nombreux cours ont un lien avec la prise en charge ambulatoire :

- Les urgences vraies ou ressenties (3h)
- Le premier recours : prise de décision dans l'incertitude (2h)
- Continuité / coordination des soins, réseaux (3h)
- Le dépistage et la prévention, la santé publique, éducation thérapeutique/conseil en santé (6h)

La proportion de cours de médecine générale par rapport au total des cours dans le cursus de TCEM à Poitier est d'environ 15,7%.

4.1.16 Personnes âgées

Le thème de la personne âgée est en diminution de 2015 à 2021 dans les thèses de médecine générale. Plusieurs hypothèses pourraient expliquer cela. Les personnes âgées une probabilité plus haute d'être hospitalisées(68), donc si le thème de l'hospitalisation est moins traité, le recrutement de personnes âgées pourrait être aussi diminué en conséquence. De plus, le recours de plus en plus

fréquent aux moyens numériques pour collecter des données pourraient exclure certaines personnes âgées des processus de recrutement(115,116). On observe une diminution globale du terme patient au profit du terme médecin, faisant suspecter que les études portent dorénavant plus sur les médecins que les patients. Une explication possible est que la réglementation est plus complexe pour les futurs thésards. Depuis la loi Jardé du 5 mars 2012, il est souvent nécessaire de faire une démarche auprès du comité de protection des personnes, de l'institut national des données de santé ou de la commission nationale de l'informatique et des libertés (117) lorsque la recherche implique au moins une personne humaine. Il est aussi possible aussi que cette tendance soit provisoire et dû à la crise du COVID 19, ayant fortement diminué l'accès direct aux patients. Une étude complémentaire sur le sujet serait nécessaire pour essayer de déterminer les raisons de cette diminution.

En DFASM et TCEM ce thème représente respectivement 4 heures soit environ 0,67% et 3 heures soit environ 3,37% de la totalité des heures de cours.

4.1.17 Domicile

Il existe un certain nombre de raisons pour lesquelles le service de soins à domicile d'un médecin généraliste peut diminuer, et dans les faits cette modalité de consultation diminue de plus en plus(71). Le coût des soins à domicile peut être un facteur, en effet, le prix du carburant augmente régulièrement, par exemple le prix du gazole a augmenté de 125.32% en 20 ans(118). La rémunération des médecins dispensant ces soins à quant à elle a augmenté de 75%(71) . L'efficacité des soins à domicile par rapport aux autres options de soins peut également être un facteur, en effet il faut compter globalement deux fois plus de temps pour une visite à domicile(71) que pour une consultation au cabinet. Les besoins du patient et son accès au transport personnel ou en commun ont aussi pu évoluer pour qu'il puisse se rendre aux rendez-vous médicaux.

Il n'existe pas de formation spécifique concernant ce type de prise en charge en DFASM ou TCEM.

4.1.18 Diabète

Cette baisse peut paraître assez paradoxale car de plus en plus de patients sont atteints de cette maladie chronique. Cependant selon Santé Publique France, les chiffres de 2020 montrent une diminution de certaines complications liées au diabète avec cependant une baisse de la qualité de certains indicateurs de suivi du diabète, qui pourraient être imputés au COVID 19 et à la plus grande difficulté d'accès aux soins pendant cette période(119). Cette baisse serait donc expliquée par une meilleure prise en charge du diabète et une meilleure prévention de l'apparition des complications.

Le diabète est enseigné à hauteur de neuf heures de cours en DFASM soit environ 1,5 % de la totalité du cursus de DFASM.

4.1.19 Anticoagulants

Nous avons pu noter une diminution de la représentation du thème des anticoagulants dans les résumés de thèses de médecine générale entre 2015 et 2021. Cela pourrait témoigner d'une meilleure connaissance des anticoagulants, en effet, les nouveaux anticoagulants oraux, ou anticoagulants oraux directs ont été maintenant mis sur le marché il y a plus de 10 ans :

- Le dabigatran (commercialisée sous le nom de Pradaxa) première AMM en 2008(120)
- Le rivaroxaban (commercialisé sous le nom de Xarelto) première AMM en 2008 (121)
- L'apixaban (commercialisé sous le nom de Eliquis) première AMM en 2011 (122)

Nous pouvons donc supposer que les études les concernant sont de moins en moins nombreuses car leurs caractéristiques sont mieux connues. Au niveau pédagogique, il y a 2 heures de cours prévues à cet effet dans le cursus de DFASM.

4.2 Validité interne

4.2.1 Avantages de cette étude

4.2.1.1 Taille de l'échantillon

Avec un effectif de 11 729 fiches de thèses extraites de la base de données Sudoc et 11 026 fiches de thèses retenues sur la période de 2015 à 2021, l'échantillon a une représentativité très importante de la population étudiée. Nous avons estimé que cela représente environ 45% du nombre total de thèse sur cette période.

4.2.1.2 Extensibilité

Cette étude présente déjà une représentativité importante, mais il est possible d'augmenter celle-ci, sans avoir besoin de recommencer l'extraction des données depuis le début. En effet nous ne nous sommes servis que d'une seule base de données, Sudoc, qui est celle qui contenait le plus d'informations. Il est possible de rajouter des résumés de thèses issus d'autres bases de données, centrales ou locales. Le seul inconvénient serait alors d'identifier les doublons, ce qui est facilement faisable grâce à l'identifiant nationale de thèse. Il sera cependant nécessaire d'adapter les méthodes d'extraction et mise en forme des informations.

4.2.1.3 Analyse ciblée

Comme nous avons pu le voir dans cette étude, il est possible d'identifier des caractéristiques de l'échantillon et de les analyser de façon indépendante du reste de l'échantillon. Ici nous avons par exemple choisi de cibler l'année 2021, mais il est possible d'analyser n'importe quelles autres caractéristiques de l'échantillon. Des analyses plus spécifiques en sous-groupes permettraient de faire des panoramas plus fins de différents thèmes importants.

4.2.1.4 Multicentrique

Cette étude ne porte pas que sur une seule origine géographique des thèses, en effet nous avons récupéré plus de 500 résumés de thèses dans 9 villes et plus de 100 résumés de thèses dans 15 villes, permettant de réduire l'effet centre. Il y a cependant un biais de sélection de thèse car certaines

viles n'ont pas été extraites de la plateforme Sudoc. D'après cette liste des 29 facultés de médecine (123), nous avons pu identifier les villes pas ou peu représentées.

Les universités non représentées sont les suivantes :

1. UFR Sciences Médicales et Pharmaceutiques - Université de Franche-Comté
2. UFR de Médecine et Sciences de la Santé - Université de Bretagne Occidentale
3. UFR de Médecine - Université de Caen Basse-Normandie
4. Faculté de Médecine - Université d'Auvergne
5. Faculté de Médecine - Université des Antilles et de la Guyane (Guadeloupe)
6. Faculté de Médecine - Université de Montpellier
7. UFR Sciences Médicales - Université de Rennes I

Les universités dont le nombre de thèses sont inférieures à 100 sont les suivantes :

1. UFR des Sciences Médicales - Université d'Angers
2. Faculté de Médecine - Université de Bourgogne
3. Faculté de Médecine Lyon Est - Université Lyon 1
4. Faculté de Médecine Lyon Sud - Université Charles Mérieux
5. Faculté de Médecine - Université de Lorraine
6. Faculté de Médecine - Université Nice Sophia-Antipolis
7. Faculté de Médecine de Tours - Université François Rabelais

Il y a donc 14 villes qui ne sont pas ou peu représentées. Cela est dû à la méthodologie employée pour rechercher les thèses dans la base de données. En effet, afin de réduire le risque d'analyser des thèses en dehors de la médecine générale, nous avons recherché spécifiquement les thèses qui mentionnaient dans le champ type de thèse « médecine générale ». Or certaines thèses de médecine générale sont indexées juste sous le terme « médecine », mais comme d'autres thèses de spécialités médicales utilisent aussi cette balise, cela aurait mené à avoir plus de thèses non-médecine générale dans l'analyse finale.

Une analyse de la répartition des principaux sujets par villes permettrait de mettre en évidence la présence ou non d'une hétérogénéité géographique. Les sujets de thèse étant forcément impactés par les thématiques locales de recherche.

4.2.1.5 *Reproductibilité*

Cette étude utilise un algorithme pour classer et analyser les données. Le script d'analyse peut être partagé et reproduit par une autre personne, les données étant en libre accès. En dehors des initiations aléatoires de certains algorithmes, les résultats sont reproductibles. Cette approche reproductible permet de modifier l'échantillon de départ, de poursuivre les analyses et de faire évoluer l'algorithme par améliorations itératives.

Tant que les données dans Sudoc sont renseignées de la même manière, l'étude peut être répétée dans le temps.

4.2.1 Désavantages de cette étude

4.2.1.1 *Biais*

4.2.1.1.1 Biais de sélection

L'indexation des thèses de médecine générale au sein de la base de données Sudoc comporte certaines erreurs et imprécisions. Notamment au niveau du type de thèse, qui en fonction de leur origine géographique sera plus ou moins bien précisé. Pour limiter ce biais lors de l'extraction des données nous avons effectué des recherches avancées dans la base de données pour éviter au maximum l'introduction de thèses non issues de la médecine générale.

4.2.1.1.2 Biais de classement

La méthodologie utilisée pour cette étude peut induire des erreurs de classement non-différentielles causées par de potentielles erreurs dans l'algorithme. L'utilisation des fonctions regex permet de réduire ce biais en autorisant une souplesse de la détection des mots, mais ne permet pas d'affirmer que la détection sera faite dans tous les cas. Cela implique qu'il y a une tendance à sous-estimer la force d'association entre deux termes.

4.2.1.1.3 Biais d'évaluation

Cette étude se repose en partie sur une interprétation subjective lors de la première partie d'analyse. En effet la liste des mots vides de sens et la lemmatisation sont faites de façon subjective. En revanche ce biais est réduit en deuxième partie lors de l'utilisation de l'apprentissage automatique, vu que l'on va se baser sur le corpus original en utilisant une autre méthode de simplification qui est la racinisation. Méthode qui a tendance à être peu parcimonieuse et peut engendrer de la confusion entre des termes de sens différents. Seules des approches plus complexes d'enrichissement sémantique permettraient de désambiguïser les termes présents dans les textes, notamment par la prise en compte du contexte.

4.2.1.1.4 Biais de mesure

Étant donné qu'un même échantillon produira les mêmes résultats en utilisant cet algorithme sur un même ordinateur, il n'y pas de biais de mesure.

4.2.1.1.5 Biais de représentativité

La période à laquelle nous nous sommes intéressés va de 2015 à 2021. Or les années 2020 et 2021 sont particulières au sens que la pandémie de Covid a drastiquement impacté les thèmes choisis par les internes en médecine. La période est donc potentiellement moins représentative du fait de cette crise sanitaire mondiale en fin de période d'étude.

4.2.1.2 *Complexité*

4.2.1.2.1 Choix des mots vides de sens

Nous pouvons observer que des mots tels que « matériel », « méthode » ou « introduction » apparaissent dans certains résultats. Cela est dû à une insuffisance de pré traitement du texte d'origine. Nous aurions pu rétrospectivement supprimer ces termes de l'analyse, au vu du peu d'informations qu'ils peuvent apporter, ainsi que tous les autres termes pouvant s'apparenter avec des mots vides de sens.

4.2.1.2.2 Choix des termes utilisées pour la recherche dans la base de données

Les termes utilisés pour rechercher les thèses dans la base de données peuvent faire varier le nombre de thèse obtenue et leur origine géographique. En effet les thèses ne sont pas indexées de manière uniforme dans la base de données. Il est donc complexe d'avoir un bon compromis entre les thèses écartées alors qu'elles sont des thèses de médecine générale d'un côté, et de l'autre côté les thèses incluses alors qu'elles ne sont pas des thèses de médecine générale.

Ici nous avons choisi de réduire au maximum la probabilité qu'une thèse non-médecine générale se retrouve dans l'analyse et malgré cela l'échantillon reste très représentatif. Or étant donné que les thèses de médecine générale sont majoritaires en médecine il est licite de se poser la question si une recherche moins stricte aurait mené aux mêmes résultats.

4.3 Validité externe

Le nombre de travaux s'intéressants aux thèmes majoritaires des thèses de médecine générale est assez faible, nous avons pu identifier ces travaux (8,11,42-44,124-127).

Notre étude est la seule utilisant des méthodes de fouille de texte, il n'a pas été possible de trouver une étude similaire.

4.3.1 Pertinence du choix de la base de données

Parmi les travaux similaires à celui-ci, la majorité (8,11,42,44,124-126) ont utilisé Sudoc au moins en partie pour extraire les thèses. Cela conforte le choix que nous avons pu faire d'utiliser cette base de données. Les autres sources de données utilisées dans la bibliographie étaient des sources locales, ou d'autres bases de données comme Docthèses (42).

4.3.2 Période d'étude

Nous avons fait le choix de porter la période d'étude de 2015 à 2021, cela par soucis d'avoir un échantillon homogène. En effet car plus les années sont antérieures à 2015, moins de thèses sont

indexées avec le résumé dans la base de données Sudoc. Donc notre intervalle d'étude est de 7 ans, soit un an de plus que la durée du premier et deuxième cycle des études médicales. Les périodes d'études retrouvées dans la littérature qui n'excèdent pas 7 ans sont au nombre de 4 (24,43,44,124), les autres peuvent aller jusqu'à une période d'étude de 12 ans (8,11,42,125,127).

4.3.3 Origine géographique des thèses

Parmi les travaux étudiés, il n'y a qu'une seule thèse ayant aussi fait le choix de prendre plusieurs origines géographiques pour la collecte de ses données : « *quatre facultés de médecine (Brest, Caen, Nantes et Rennes)* » (42). Cependant cela reste cantonné au nord-ouest de la France. Le désavantage est qu'il y aura bien plus d'effet centre parmi ces études comme mentionné dans cette thèse strasbourgeoise : « *il existe un groupe de travail collaboratif entre le DMG et le service d'infectiologie, ce qui peut expliquer en partie la grande implication des étudiants dans l'infectiologie (10,5% des thèses au total mentionnent ce sujet)* » (24). Effectivement à l'échelle d'une faculté, si un département est particulièrement prolifique en accompagnement de thèses de médecine générale, il est probable que cela influera les résultats.

4.3.4 Effectifs

Aucun des travaux étudiés ne dépasse les 800 thèses analysées (8,11,42–44,124–127), cependant à cause de la grande disparité des périodes d'études, il vaut mieux raisonner en terme de nombre moyen de thèse par année. Le nombre de thèses en moyenne étudié le plus bas est de 28 thèses par années (42) jusqu'à un maximum de 134 thèses par années étudiées (24,44). Cette étude comporte donc l'effectif le plus important avec 11 026 thèses analysées et 1575 thèses par années en moyenne.

4.3.5 Tri des sujets appartenant à la médecine générale ou non

Dans ce travail nous ne nous sommes pas intéressés au fait que les sujets des thèses étudiées soient dans le champ de la médecine générale ou pas. Les résumés de thèses ont été exclusivement sélectionnés s'ils étaient présentés pour le DES de médecine générale. Dans l'intégralité des

précédents travaux sur le sujet il y a eu un classement des thèses de médecine générale selon que le sujet s'accordait à la définition de la World Organization of National Colleges, Academies and Academic Associations of General Practitioners (WONCA) (128) ou non (8,11,42–44,124–127).

4.3.6 Taxonomie des sujets

Certains travaux ont cherché à retrouver aussi les thèmes majoritaires dans les thèses (8,11,42–44,125,126). Pour cela elles ont effectué des tris d'après des grilles de classification :

- Medical Subject Headings (MeSH) (129) et Catalogage et l'Indexation des Sites Médicaux de langue Française (CISMeF) (129) pour les travaux suivants (8,42,43).
- International Classification of Primary Care (ICPC-2) (130) de la WONCA ou Classification internationale des soins primaires (CISP) pour les travaux suivants (8,11,24,44)

Pour cette étude, nos travaux préliminaires nous ont permis de suspecter que l'on perdait de l'information en utilisant un système de classification, ce que divers travaux suggèrent aussi « *La majorité des mots clés utilisés par les thésards pour caractériser leur travail ne faisait pas parti des mots clés indexés selon la référence du MesH* » (8), et qu'ils ne sont pas adaptés à toutes les situations (131,132). L'objectif de notre étude étant d'identifier des thématiques spécifiques et non générales, nous n'avons pas utilisé ce type de système de classification. La classification des données a été effectuée par apprentissage automatique. L'utilisation d'ontologies médicales pour indexer les termes présents dans les thèses aurait néanmoins permis de se limiter aux thématiques médicales et enlever des potentielles thématiques liées à la méthodologie ou la population d'étude.

4.3.7 Évolution des thématiques

Sur l'ensemble des travaux, quelques-uns ont pu traiter l'évolution de certains thèmes (43,44,124).

Genin et Agrifolio ne s'intéressent qu'à l'évolution de la proportion médecin-patient dans les thèses :

« Dans notre thèse [...] les travaux MG sont moins centrées patient jusqu'en 2013. Puis cette tendance s'est inversée, si bien que les thèses soutenues en 2014 et 2015 étaient plus centrées patient » (124).

LEONARD Mathieu s'intéresse lui aussi à l'évolution de ces thématiques : *« trois catégories qui se détachent : général et non spécifié (21% des thèses), procédures (11%) et aucun (13%). [...] Les chiffres sont similaires pour la période de 2008 à 2010 avec près de 40% pour les trois catégories les plus générales » (44)*

Balva et Al (43) effectuent une comparaison de l'évolution des thématiques retrouvées sur deux périodes de 1997 à 1999 et de 2007 à 2009, avec un effectif de 43 thèses par année en moyenne, pouvant expliquer les fortes variations observées. Cette dernière reste l'étude la plus proche de la nôtre pour effectuer des comparaisons sur les évolutions des sujets.

4.3.8 Comparaison des résultats

4.3.8.1 Âge moyen

Une des données systématiquement recueillies est l'âge au moment de la soutenance de thèse.

De 1997 à 1999 nous avons une valeur de 31,9 ans chez Balva et al (43), de 2004 à 2014 ; Chiaro retrouve une valeur de 30,7 ans (8), et nous retrouvons dans notre étude une moyenne à 30 ans pour la période 2015-2021. Les autres auteurs retrouvent des valeurs cohérentes qui diminuent depuis 1997 (11,24,44,124). Il semblerait donc au vu de ces résultats que l'âge moyen de la soutenance de thèse diminue.

4.3.8.2 Sujets prépondérants

Genin et Agrifolio (124) observent que les thèses soutenues à partir de 2013 sont plus centrées sur le patient. Nous retrouvons au contraire dans notre étude une évolution inverse depuis 2015 avec

le terme médecin prépondérant et en augmentation par rapport au terme patient. Il est à noter qu'en 2015 les thèses mentionnent quasiment à part égale les termes patients et médecins, ce qui serait cohérent avec ce qu'affirmait Genin et Agrifolio (124).

Concernant les thèmes prépondérants nous retrouvons ces 4 thèmes les plus probables :

- Formation des médecins généralistes (3,7%)
- Pratiques en médecine générale (3,4%)
- Patients hospitalisés (3,2%)
- Relation médecin-patient (3,2%)

Labonde et al (125) trouvent de 2004 à 2016 ce classement en utilisant les mots clés d'indexation :

- Médecine générale (21%)
- Urgences (2,3%)
- Médecins généralistes (1,7%)
- Prévention (1,4%)
- Enfants et personnes âgées (1,3%)

Nous allons ignorer les résultats concernant l'item médecine générale et médecins généralistes. Nous retrouvons dans notre thème « recours aux urgences » une probabilité de 2,3 %, ce qui est exactement ce que l'auteur a retrouvé, au niveau de notre thème « médecine préventive » nous retrouvons 1,6% et l'auteur 1,4% ce qui est assez proche. Pour ce qui est des enfants et des personnes âgées nous trouvons une valeur plus élevée avec respectivement 1,7% et 1,8 % contre 1,3% pour Labonde et al.

Vernaudon Damien (24) lui, a classé de manière thématique des thèses strasbourgeoises de 2016 à 2017.

Il trouve les résultats suivants :

- Pratique / formation médicale (N=66, 15,87%)

- Gériatrie (N=49, 11,79%)
- Thérapeutique (N=43, 10,34%).

La comparaison avec cet auteur reste assez épineuse vu que le classement choisi est très général. Pour avoir des thèmes comparables nous devons donc fusionner nos thèmes avec ceux qui peuvent correspondre au thème pratique / formation médicale :

- Formation des médecins généralistes 3,7%
- Pratiques en médecine générale 3,4 %
- Relation médecin-patient 3,2 %
- Outils diagnostiques 2,4 %
- Formation des internes 2,1 %
- Médecine préventive 1,6 %

Nous arrivons à un total de 16,4% pour ce thème composite, assez proche des résultats observés par l'auteur.

Pour le thème gériatrie nous allons nous livrer au même exercice :

- Patients hospitalisés 3,2 %
- Recours aux urgences 2,3 %
- Personne âgée fragile 1,8 %
- Soins à domicile et fin de vie 1,7%

Soit 9 % au total, soit un peu moins que les 11,8% observés par l'auteur, bien qu'il cite dans sa thèse qu'il y a une forte activité de la gériatrie dans la direction des thèses au CHU de Strasbourg.

Pour le thème thérapeutique nous allons à nouveau nous livrer au même exercice :

- Pratiques de prescription 1,8%
- Troubles du sommeil et psychiatrique 1,8%
- Suivi gynécologique et contraception chez la femme 1,8%

- Prise en charge des douleurs chroniques 1,8%
- Diabète 1,7%
- Vaccination 1,7%
- Sevrage alcool et tabac 1,5 %

Soit 12,1% contre les 10,3% retrouvés par l'auteur.

Les résultats de cet auteur sont cohérents avec ce que l'on peut retrouver dans notre travail.

Pour Balva et al., l'auteur a observé deux périodes distinctes de 1997 à 1999 et de 2007 à 2009 puis les a classés en deux périodes respectivement P1 et P2. Il trouve les résultats suivants :

« En P1, les disciplines dominantes étaient l'oncologie, la pathologie infectieuse, la psychiatrie et la cardiologie. En P2, les disciplines dominantes étaient l'exercice, la gynécologie et la médecine d'urgence » (43).

Nos résultats semblent foncièrement différents pour ce qui est de la période P1. Au sein de P2 nous pouvons retrouver des sujets se rapprochant de ce qu'évoque l'auteur. Pour l'exercice qu'il évoque nous pouvons le comparer au sujet « pratiques en médecine générale » (3,2%) ou installation et exercice de la médecine générale en milieu rural (2,5%). Pour la médecine d'urgence nous pouvons le comparer plus aisément au sujet du recours aux urgences (2,3%).

4.3.8.3 *Évolution des sujets*

Balva et al a remarqué une diminution de la population adulte malade dans les études de 49 à 37% et une augmentation de la population des études concernant les médecins de 9 à 20% entre les deux périodes étudiées. Nous avons nous aussi pu noter une diminution du terme patient au sein du corpus ainsi qu'une augmentation du terme médecin. Cette inversion des courbes peut être dûe à la crise démographique médicale actuelle qui ferait qu'on se pose plus de questions sur l'exercice du médecin. Ou alors au fait qu'il soit plus simple administrativement et en pratique de réaliser des études

chez les médecins plutôt que chez les patients., les études sur questionnaire étant notamment plus encadrées depuis la loi Jardé.

Il a aussi noté une augmentation des thèses à orientation « dépistage » de 1,5 à 9 % entre 1997-1999 et 2007-2009. Le dépistage faisant parti des actes de prévention nous pouvons y opposer les résultats trouvés en Figure 39, page 106 sur la prévention. Nous trouvons une augmentation de 14 à 16,7 % soit une croissance de 16,7% sur la période 2015-2021.

5 Conclusion

Cette étude explore l'évolution des thématiques des thèses de médecine générale en France en 2021 et analyse l'évolution de ceux-ci de 2015 à 2021. La méthodologie développée à base de fouille automatisée de texte permet d'explorer l'ensemble des thèses disponibles sur la plateforme Sudoc en analysant automatiquement leur résumé.

Les thèmes prépondérants en 2021 portent sur la formation et les pratiques des médecins généralistes, les patients hospitalisés et la relation médecin-patient.

Les thématiques ayant la plus forte augmentation sur la période 2015-2020 concernent les risques pandémiques ainsi que les outils informatiques. Ces augmentations font sens dans le contexte sanitaire actuel où l'épidémie de COVID qui a commencé en 2020 a développé de façon importante l'usage de la téléconsultation. À l'opposé, certains sujets sont de moins en moins populaires, comme la question des anticoagulants en médecine générale ou de la personne âgée.

Les thèmes qui restent prépondérants de 2015 à 2021 traitent de la formation et des pratiques des médecins généralistes, de la relation médecin-patient, de la médecine préventive, des prises en charges ainsi que de l'installation et de l'exercice de la médecine générale.

Se basant sur des techniques automatisées de science des données, cette analyse peut être réactualisée annuellement. Elle pourrait être la base d'un outil permettant de guider les futurs travaux de thèse. En supposant que les internes choisissent leur sujet de thèse en fonction des potentiels manques de connaissance qu'ils ont, cette analyse permet également de mettre en évidence des thématiques importantes à approfondir pendant le cursus médical.

VU et approuvé
Strasbourg, le 11 JAN. 2023
Le Doyen de la Faculté de
Médecine, Médecine et Sciences de la Santé
Professeur Jean SPINIA

VU
Strasbourg, le 3 Janvier 2023
Le président du jury de thèse
Professeur Thierry PELACCA

6 Annexes

Annexe 1 : Liste des termes changés pour la lemmatisation du corpus

actes->acte	études->étude	physiques->physique
actifs->actif	étudiants->étudiant	plaies->plaie
activités->activité	examens->examen	populations->population
addiction->addictologie	facteurs->facteur	pratiques->pratique
admis ->admission	familles->famille	précaires->précarité
adolescents->adolescent	favorables->favorable	prescriptions->prescription
âgée->âgé	femmes->femme	prescrire->prescription
âgées->âgé	filières->filière	prescrit->prescription
âgés->âgé	formes->forme	prescrits->prescription
aidants->aidant	fractures->fracture	prescrivaient->prescription
ambulatoires->ambulatoire	freins->frein	principaux->principale
années->année	fumeurs->tabagisme	professionnels->professionnel
ans -> âge	généralistes->généraliste	programmes->programme
antalgiques->antalgique	gériatrique->gériatrie	psychiatrique->psychiatrie
antécédents->antécédent	gériatriques->gériatrique	psychiatriques->psychiatrique
antibiotique->antibiothérapie	gestes->geste	psychologiques->psychologique
antibiotiques->antibiotique	grossesses->grossesse	questionnaires->questionnaire
anticipées->anticipée	groupes->groupe	questions->question
appels->appel	gynécologique->gynécologie	recommandations->recommandation
biologiques->biologique	gynécologiques->gynécologique	relationnelles->relationnelle
cabinets->cabinet	gynécologues->gynécologue	repérer->repérage
campagnes->campagne	hommes->homme	réponse->répondre
cancers->cancer	horaires->horaire	réponses->réponse
cancers->cancer	hospitalisations->hospitalisation	résidents->résident
cardiovasculaires->cardiovasculaire	hospitalisés->hospitalisation	respiratoires->respiratoire
catégories->catégorie	ide -> infirmier	résultats->résultat
chroniques->chronique	identifiés->identifier	rôles->rôle
chutes ->chute	indésirables->indésirable	sages->sage
cliniques->clinique	infections->infection	satisfaits->satisfaction
comorbidités->comorbidité	infirmière->infirmier	scores->score
compétences->compétence	infirmières->infirmier	secondaires->secondaire
complémentaires->complémentaire	infirmiers->infirmier	services->service
complications->complication	internes->interne	seules->seule
conjugales->conjugale	interrogés->interrogé	sévère->sévérité
connaissaient= connaissance	interventions->intervention	sévères->sévère
connaissances->connaissance	jeunes->jeune	sexuel ->sexualité
conséquences->conséquence	jours->jour	sexuelle ->sexualité
consommations->consommation	kinésithérapeutes->kinésithérapeute	sexuelles ->sexualité
consultations->consultation	lésions->lésion	sexuels ->sexualité

contraceptif->contraception	maladies->maladie	signes->signe
contraceptifs->contraception	médecins->médecin	sociales->social
contraceptions->contraception	médicamenteuse->médicament	spécialisée->spécialisé
contraceptive->contraception	médicamenteuses->médicament	spécialisés ->spécialisé
contraceptives->contraception	médicamenteux->médicament	sportif->sport
couples->couple	médicaments->médicament	sportifs->sport
décédé->décès	médicaux->médical	sportive->sport
décédés->décès	mères->mère	sports->sport
dépister->dépistage	méthodes->méthode	stages->stage
dépressif->dépression	mort ->décès	substances->substance
diabétique->diabète	nourrissons->nourrisson	sujets->sujet
diabétiques->diabète	palliatifs->palliatif	tabac->tabagisme
diagnostique ->diagnostic	palliative->palliatif	tabagique->tabagisme
diagnostiqué ->diagnostic	parent ->parentalité	téléphoniques->téléphoniques
diagnostiquée ->diagnostic	parents->parent	thèses->thèse
diagnostiques ->diagnostic	partenaires->partenaire	traitements->traitement
difficultés->difficulté	pathologies->pathologie	traumatismes->traumatisme
directives->directive	patientes->patiente	troubles->trouble
dossiers->dossier	patients->patient	universitaire->université
douleurs->douleur	pédiatrique->pédiatrie	universitaires->université
droits->droit	pédiatriques->pédiatrique	urgences->urgence
écrans->écran	perceptions ->perception	urgentistes->urgentiste
effets->effet	perçoivent->perception	utérin->utérus
élèves->élève	pères->père	vaccin ->vaccination
émotionnelle->émotionnel	personnes->personne	vaccinal.->vaccination
enceintes->enceinte	pharmacie ->pharmacien	vaccinale.->vaccination
enfants->enfant	pharmaciens->pharmacien	vaccinales ->vaccination
entretiens->entretien	physiques->physique	vaccinations ->vaccination
épidémique->épidémie	plaies->plaie	vaccins ->vaccination
équipes ->équipe	populations->population	victimes->victime

Annexe 2 : Liste des mots vides de sens appliqués à l'ensemble de l'analyse

"dissertation", "a", "abord", "absolument", "afin", "ah", "ai", "aie", "aient", "aies", "ailleurs", "ainsi", "ait", "allaient", "allo", "allons", "allô", "alors", "antérieur", "antérieure", "antérieures", "après", "assez", "attendu", "au", "aucun", "aucune", "aucuns", "aujourd", "aujourd'hui", "aupres", "auquel", "aura", "aurai", "auraient", "aurais", "aurait", "auras", "aurez", "auriez", "aurions", "aurons", "auront", "aussi", "autre", "autrefois", "autrement", "autres", "autrui", "aux", "auxquelles", "auxquels", "avaient", "avais", "avait", "avant", "avec", "avez", "aviez", "avions", "avoir", "avons", "ayant", "ayez", "ayons", "b", "bah", "bas", "basee", "bat", "beau", "beaucoup", "bien", "bigre", "bon", "boum", "bravo", "brrr", "c", "car", "ce", "ceci", "cela", "celle", "celle-ci", "celle-là", "celles", "celles-ci", "celles-là", "celui", "celui-ci", "celui-là", "celà", "cent", "cependant", "certain", "certaine", "certaines", "certains", "certes", "ces", "cet", "cette", "ceux", "ceux-ci", "ceux-là", "chacun", "chacune", "chaque", "cher", "chers", "chez", "chiche", "chut", "chère", "chères", "ci", "cinq", "cinquante", "cinquantième", "cinquième", "clac", "clic", "combien", "comme", "comment", "comparable", "comparables", "compris", "concernant", "contre", "couic", "crac", "d", "da", "dans", "de", "debout", "dedans", "dehors", "deja", "delà", "depuis", "dernier", "derniere", "derrière", "des", "desormais", "desquelles", "desquels", "dessous", "dessus", "deux", "deuxième", "deuxièmement", "devant", "devers", "devra", "devrait", "different", "différentes", "différents", "différent", "différente", "différentes", "différents", "dire", "directe", "directement", "dit", "dite", "dits", "divers", "diverse", "diverses", "dix", "dix-huit", "dix-neuf", "dix-sept", "dixième", "doit", "doivent", "donc", "dont", "dos", "douze", "douzième", "dring", "droite", "du", "duquel", "durant", "dès", "début", "désormais", "e", "effet", "egale", "egalement", "egales", "eh", "elle", "elle-même", "elles", "elles-mêmes", "en", "encore", "enfin", "entre", "envers", "environ", "es", "essai", "est", "et", "étant", "etc", "etre", "eu", "eue", "eues", "euh", "eurent", "eus", "eusse", "eussent", "eusses", "eussiez", "eussions", "eut", "eux", "eux-mêmes", "exactement", "excepté", "extenso", "extérieur", "eûmes", "eût", "eûtes", "f", "fais", "faisaient", "faisant", "fait", "faites", "façon", "feront", "fi", "flac", "floc", "fois", "font", "force", "furent", "fus", "fusse", "fusses", "fussiez", "fussions", "fut", "fûmes", "fût", "fûtes", "g", "gens", "h", "ha", "haut", "hein", "hem", "hep", "hi", "ho", "holà", "hop", "hormis", "hors", "hou", "houp", "hue", "hui", "huit", "huitième", "hum", "hurrah", "hé", "hélas", "i", "ici", "il", "ils", "importe", "j", "je", "jusqu", "jusque", "juste", "k", "l", "la", "laisser", "laquelle", "las", "le", "lequel", "les", "lesquelles", "lesquels", "leur", "leurs", "longtemps", "lors", "lorsque", "lui", "lui-même", "lui-même", "là", "lès", "m", "ma", "maint", "maintenant", "mais", "malgre", "malgré", "maximale", "me", "meme", "memes", "merci", "mes", "mien", "mienne", "miennes", "miens", "mille", "mince", "mine", "minimale", "moi", "moi-meme", "moi-même", "moindres", "moins", "mon", "mot", "moyennant", "multiple", "multiples", "même", "mêmes", "n", "na", "naturel", "naturelle", "naturelles", "ne", "neanmoins", "nécessaire", "nécessairement", "neuf", "neuvième", "ni", "nombreuses", "nombreux", "nommés", "non", "nos", "notamment", "notre", "nous", "nous-mêmes", "nouveau", "nouveaux", "nul", "néanmoins", "nôtre", "nôtres", "o", "oh", "ohé", "ollé", "olé", "on", "ont", "onze", "onzième", "ore", "ou", "ouf", "ouias", "oust", "ouste", "outré", "ouvert", "ouverte", "ouverts", "o", "où", "p", "paf", "pan", "par", "parce", "parfois", "parle", "parlent", "parler", "parmi", "parole", "parseme", "partant", "particulier", "particulière", "particulièrement", "pas", "passé", "pendant", "pense", "permet", "personne", "personnes", "peu", "peut", "peuvent", "peux", "pff", "pfft", "pft", "pif", "pire", "pièce", "plein", "plouf", "plupart", "plus", "plusieurs", "plutôt", "possessif", "possessifs", "possible", "possibles", "pouah", "pour", "pourquoi", "pourrais", "pourrait", "pouvait", "prealable", "precisement", "premier", "première", "premièrement", "pres", "probable", "probante", "procedant", "proche", "près", "psitt", "pu", "puis", "puisque", "pur", "pure", "q", "qu", "quand", "quant", "quant-à-soi", "quanta", "quarante", "quatorze", "quatre", "quatre-vingt", "quatrième", "quatrièmement", "que", "quel", "quelconque", "quelle", "quelles", "quelqu'un", "quelque", "quelques", "quels", "qui", "quiconque", "quinze", "quoi", "quoique", "r", "rare", "rarement", "rares", "relative", "relativement", "remarquable", "rend", "rendre", "restant", "reste", "restent", "restrictif", "retour", "revoici", "revoilà", "rien", "s", "sa", "sacrebleu", "sait", "sans", "sapristi", "sauf", "se", "sein", "seize", "selon", "semblable", "semblaient", "semble", "semblent", "sent", "sept", "septième", "sera", "serai", "seraient", "serais", "serait", "seras", "serez", "seriez", "serions", "serons", "seront", "ses", "seul", "seule", "seulement", "si", "sien", "sienne", "siennes", "siens", "sinon", "six", "sixième", "soi", "soi-même", "soient", "sois", "soit", "soixante", "sommés", "son", "sont", "sous", "souvent", "soyez", "soyons", "spécifique", "spécifiques", "speculatif", "stop", "strictement", "subtiles", "suffisant", "suffisante", "suffit", "suis", "suit", "suivant", "suivante", "suivantes", "suivants", "suivre", "sujet", "superpose", "sur", "surtout", "t", "ta", "tac", "tandis", "tant", "tardive", "te", "tel", "telle", "tellement", "telles", "tels", "tenant", "tend", "tenir", "tente", "tes", "tic", "tien", "tienne", "tiennes", "tiens", "toc", "toi", "toi-même", "ton", "touchant", "toujours", "tous", "tout", "toute", "toutefois", "toutes", "treize", "trente", "tres", "trois", "troisième", "troisièmement", "trop", "très", "tsouin", "tu", "té", "u", "un", "une", "unes", "uniformement", "unique", "uniques", "uns", "v", "va", "vais", "valeur", "vas", "vers", "via", "vif", "vifs", "vingt", "vivat", "vive", "vives", "vlan", "voici", "voie", "voient", "voilà", "vont", "vos", "votre", "vous", "vous-mêmes", "vu", "vé", "votre", "vôtres", "w", "x", "y", "z", "zut", "à", "à", "ça", "ès", "étaient", "étais", "était", "étai", "étiez", "étions", "été", "éte", "éteés", "étés", "êtes", "être", "ô", "||", "tis", "twas", "ve", "10", "39", "a", "a's", "able", "ableabout", "about", "above", "abroad", "abst", "accordance", "according", "accordingly", "across", "act", "actually", "ad", "added", "adj", "adopted", "ae", "af", "affected", "affecting", "affects", "after", "afterwards", "ag", "again", "against", "ago", "ah", "ahead", "ai", "ain't", "aint", "al", "all", "allow", "allows", "almost", "alone", "along", "alongside", "already", "also", "although", "always", "am", "amid", "amidst", "among", "amongst", "amoungst", "amount", "an", "and", "announce", "another", "any", "anybody", "anyhow", "anymore", "anything", "anyway", "anysways", "anywhere", "ao", "apart", "apparently", "appear", "appreciate", "appropriate", "approximately", "aq", "ar", "are", "area", "areas", "aren", "aren't", "arent", "arise", "around", "arpa", "as", "aside", "ask", "asked", "asking", "asks", "associated", "at", "au", "auth", "available", "aw", "away", "awfully", "az", "b", "ba", "back", "backed", "backing", "backs", "backward", "backwards", "bb", "bd", "be", "became", "because", "become", "becomes", "becoming", "been", "before", "beforehand", "began", "begin", "beginning", "beginnings", "begins", "behind", "being", "beings", "believe", "below", "beside", "besides", "best", "better", "between", "beyond", "bf", "bg", "bh", "bi", "big", "bill", "billion", "biol", "bj", "bm", "bn", "bo", "both", "bottom", "br", "brief", "briefly", "bs", "bt", "but", "buy", "bv", "bw", "by", "bz", "c", "c'mon", "c's", "ca", "call", "came", "can", "can't", "cannot", "cant", "caption", "case", "cases", "cause", "causes", "cc", "cd", "certain", "certainly", "cf", "cg", "ch", "changes", "ci", "ck", "cl", "clear", "clearly", "click", "cm", "cmon", "cn", "co", "co.", "com", "come", "comes", "computer", "con", "concerning", "consequently", "consider", "considering", "contain", "containing", "contains", "copy", "corresponding", "could", "could've", "couldn", "couldn't", "couldn't", "course", "cr", "cry", "cs", "cu", "currently", "cv", "cx", "cy", "cz", "d", "dare", "darent", "darent", "date", "de", "dear", "definitely", "describe", "despite", "detail", "did", "didn't", "didn't", "didn't", "differ", "different", "differently", "directly", "dj", "dk", "dm", "do", "does", "doesn", "doesn't", "doesnt", "doing", "don", "don't", "done", "dont", "doubtful", "down", "downed", "downing", "downs", "downwards", "due", "during", "dz", "e", "each", "early", "ec", "ed", "edu", "ee", "effect", "eg", "eh", "eight", "eighty", "either", "eleven", "else", "elsewhere", "empty", "end", "ended", "ending", "ends", "enough", "entirely", "er", "es", "especially", "et", "et-al", "etc", "even", "evenly", "ever", "evermore", "every", "everybody", "everyone", "everything", "everywhere", "ex", "exactly", "example", "except", "f", "face", "faces", "fact", "facts", "fairly", "far", "farther", "felt", "few", "fewer", "ff", "fi", "fifteen", "fifth", "fifty", "fif", "fill", "find", "finds", "fire", "first", "five", "fix", "fj", "fk", "fm", "fo", "followed", "following", "follows", "for", "forever", "former", "formerly", "forth", "forty", "forward", "found", "four", "fr", "free", "from", "front", "full", "fully", "further", "furthered", "furthering", "furthermore", "furthers", "fx", "g", "ga", "gave", "gb", "gd", "ge", "general", "generally", "get", "gets", "getting", "gf", "gg", "gh", "gi", "give", "given", "gives",

Annexe 3: Liste des mots supprimés de l'analyse de corrélation entre les termes

"compte", "tenu", "quasi", "totalité", "mots", "clés", "résultats", "conclusion", "prise", "charge", "morbi", "médecine", "générale", "critère", "jugement", "resultats", "méthode", "aquitaine", "calais", "methode", "saint", "seine", "denis", "rhône", "isère", "savoie", "logistique", "retranscrits", "enregistrés", "principal", "secondaire", "statistiques", "menés", "poitiers", "bordeaux", "nvivo", "manque", "frein", "partagée", "décision", "positifs", "négatifs", "positive", "négative", "échantillonnage", "neige", "avantages", "inconvenients", "met", "évidence", "paca", "pistes", "amélioration", "salle", "attente", "loire", "vendée", "phénoménologique", "interprétative", "double", "codage", "île", "rôle", "individuels", "qualitative", "entretien", "semi", "dirigés", "littérature", "revue", "pubmed", "articles", "bases", "cochrane", "essais", "randomisés", "delphi", "théorisation", "ancrée", "différence", "significative", "médecin", "généraliste", "consensus", "experts", "rétrospective", "monocentrique", "transversale", "descriptive", "questionnaire", "tour", "ap", "apa", "soins", "primaires", "haute", "autorité", "centre", "hospitalier", "assurance", "maladie", "problème", "majeur", "autorités", "sanitaires", "enjeu", "santé", "publique", "élève", "classe", "collaboration", "interprofessionnelle", "bilan", "biologique", "strasbourg", "soin", "parcours", "facteur", "risque", "catégorie", "socio", "démographiques", "longue", "durée", "sites", "internet", "mieux", "prendre", "mail", "envoyé", "année", "dernières", "quotidien", "vivant", "comparativement", "performance", "zones", "pallier", "aide", "logiciel", "élevée", "prévalence", "prévenir", "systémique", "obtenues", "répondre", "estiment", "informés", "central", "constituent", "difficulté", "rencontrées", "heure", "actuelle", "thérapie", "domaine", "établissements", "décrits", "prévention", "primaire", "année", "dernières", "téléphone", "contactés", "spécialiste", "thérapies", "approche", "centrée", "multivariée", "associés", "forces", "armées", "département", "vienna", "faisait", "oise", "somme", "toulouse", "nantes", "grand", "public", "volume", "valeurs", "inscrit", "démarche", "hauts", "france", "explorer", "vécu", "prédictifs", "tolérance", "assurer", "continuité", "ère", "direct", "science", "alpes", "région", "pronostic", "survie", "inclusion", "cohorte", "critères", "réalisées", "analyses", "méta", "revues", "incluses", "systématique", "sélectionnées", "identifiées", "prisma", "étude", "chercheurs", "interprétation", "inspirée", "variation", "échantillonnage", "recrutement", "analyse", "thématique", "retranscrits", "suffisance", "conclure", "google", "effectué", "triangulation", "retranscrits", "suffisance", "conclure", "recueil", "données", "sélection", "recueillies", "saturation", "premières", "semaine", "heures", "ème", "répondre", "initial", "papier", "format", "rendu", "étapes", "code", "journée", "prospective", "interventionnelle", "téléphoniques", "appel", "téléphonique", "donné", "moitié", "tiers", "suivi", "régulier", "renouvellement", "ordonnance", "représentation", "image", "connu", "qualitatives", "issus", "article", "ressort", "sources", "informations", "fiches", "fiche", "créé", "évaluant", "outil", "utilisation", "utilisateurs", "scientifiques", "preuves", "issues", "enjeux", "documents", "utiles", "destination", "constitue", "phase", "leviers", "pilote", "synthèse", "constituer", "rendez", "émerger", "verbatim", "gain", "cible", "panel", "concernaient", "mesurer", "pertinence", "élaboration", "références", "travaux", "existants", "motivationnel", "habitudes", "résaux", "sociaux", "relation", "confiance", "masculin", "support", "vidéos", "message", "fréquent", "traumatisme", "lésion", "irm", "allergie", "confirmée", "alimentaire", "comportement", "repérage", "précoce", "indésirable", "molécules", "terme", "motif", "consultation", "professionnel", "personnelle", "médicale", "démographie", "douleur", "antalgique", "paracetamol", "déclaration", "obligatoire", "clinique", "examen", "complémentaire", "échographie", "mauvaise", "observance"

Annexe 4 : Liste des mots supprimés de l'analyse de corrélation entre les bigrammes

"méthode etude","essais contrôlés","auvergne rhône","seine saint","entretien individuels","questionnaire auto","relation médecin","statistiquement significative","entretien semi","étude monocentrique","étude qualitative","étude épidémiologique","analyse phénoménologique","théorisation ancrée","dirigés réalisés","observationnelle descriptive","semi dirigés","jugement principal","étude quantitative","revue systématique","données pubmed","étude prospective","étude transversale","survie globale","contrôlés randomisés","étude rétrospective","pubmed cochrane","étude révèle","diminution significative","derniers mois","étude rétrospective","analyse multivariée","google scholar","poitou charentes","polynésie française","inclus patient","individuels semi","données recueillies","descriptive transversale","patient présentaient","cohorte prospective","entretien téléphonique","publique majeur","problème majeur","étude observationnelle","quantitative réalisée","mots clés","monocentrique réalisée","ancrée résultat","théorie ancrée ","enjeu majeur","polynésie française","approche centrée","observationnelle transversale","descriptive observationnelle","revue narrative","étude observationnelles","épidémiologique observationnelle","étude incluses","critères secondaire","étude montre","méthode enquête","enquête qualitative","quantitative transversale","descriptive quantitative","association statistique","significativement associés","facteur liés","méta analyse","appel téléphonique","lien significatif","significativement associés","étude vise ","critères secondaire","méthode revue","étude montrent","descriptive rétrospective","pratique clinique","médecine générale","semi dirigé","examen complémentaire","observationnelle prospective","questionnaire adressé","théorie ancrée","saint étienne","mieux comprendre","descriptive monocentrique","prospective observationnelle","quasi totalité","charge initiale","étude vise")

374 *Essai sur les Maladies*

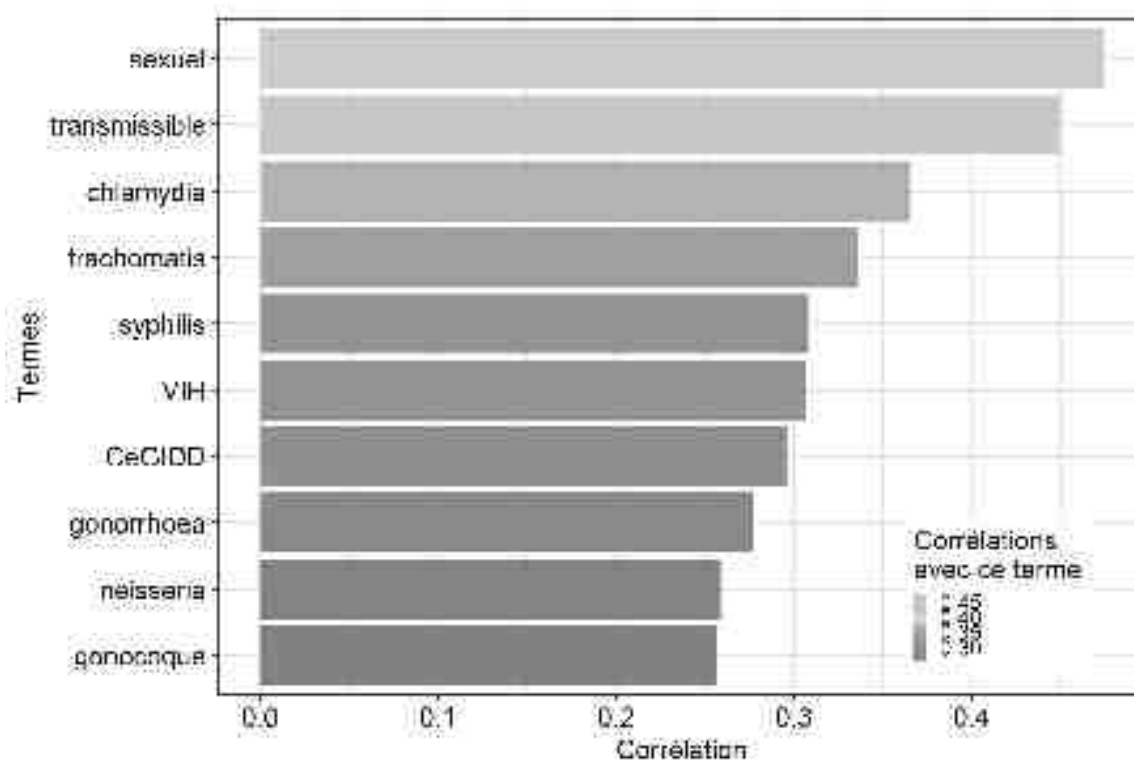
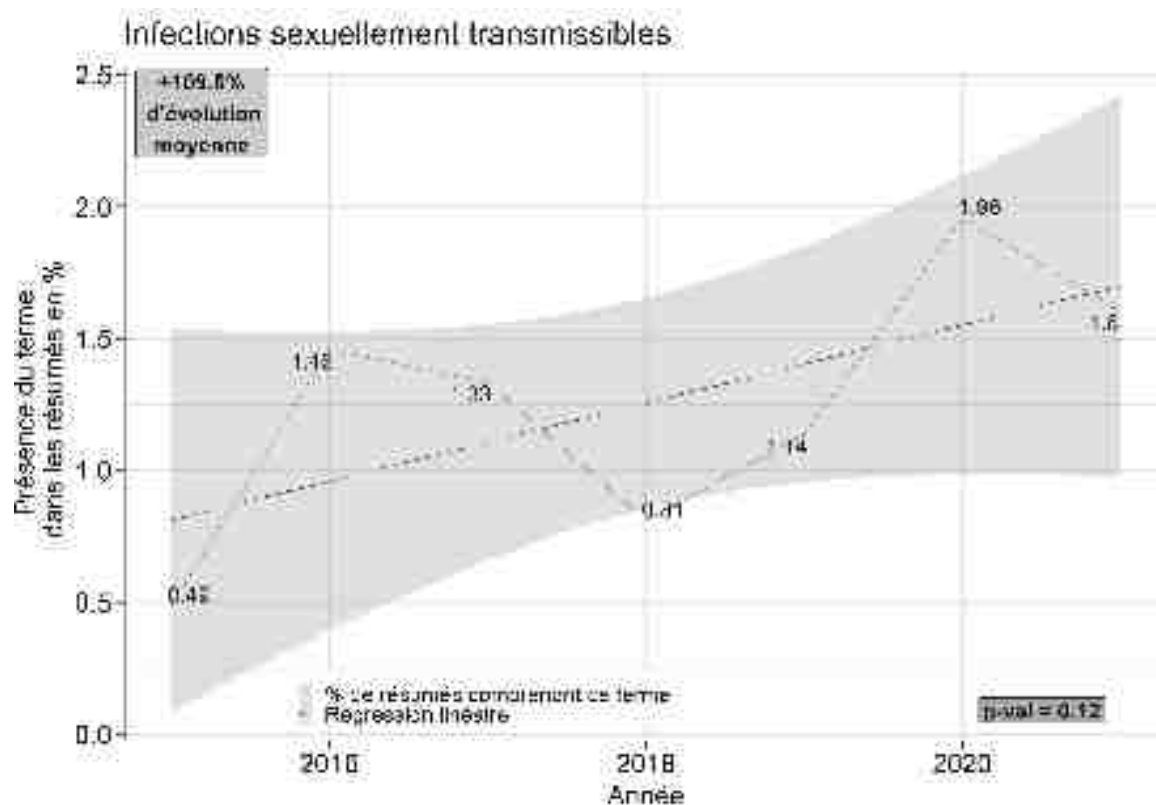
le nez, & éprouva un flux modéré par le ventre.

La vie sédentaire que menent ces Ouvriers, & les Tailleurs sur-tout, produit donc une disposition vicieuse dans les organes, & une abondance nuisible d'humeurs. Il n'en est pas de même des autres Artisans qui, quoique assis, exercent leurs bras, leurs pieds, & tout leur corps à leur ouvrage, comme les Potiers de terre, les Tisserands, &c. : les mouvemens qu'ils exécutent, chassent les impuretés de leur sang, & rendent leur santé plus robuste. La douleur des reins est encore une maladie particulière aux Ouvriers assis. Ainsi Plaute a dit : « Les » lombes deviennent douloureuses lorsqu'on reste assis trop long-temps, ainsi » que les yeux par des regards trop » fixes (a) ».

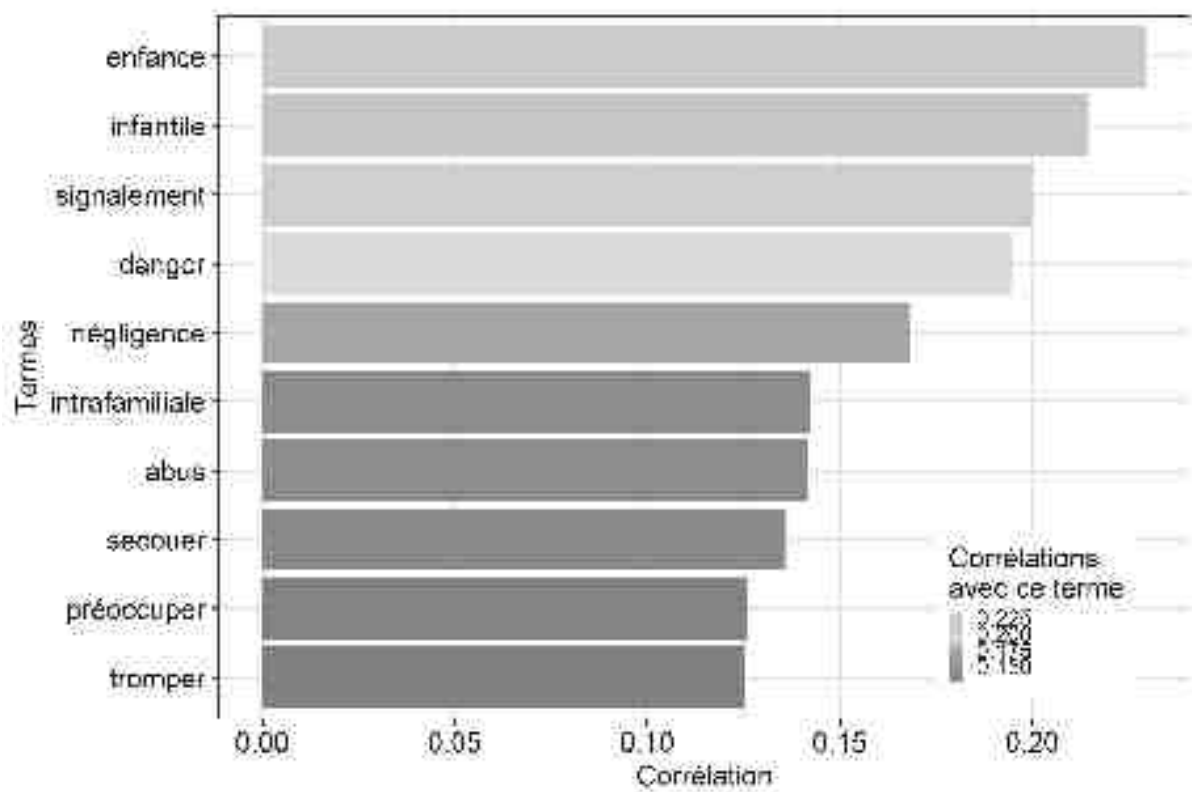
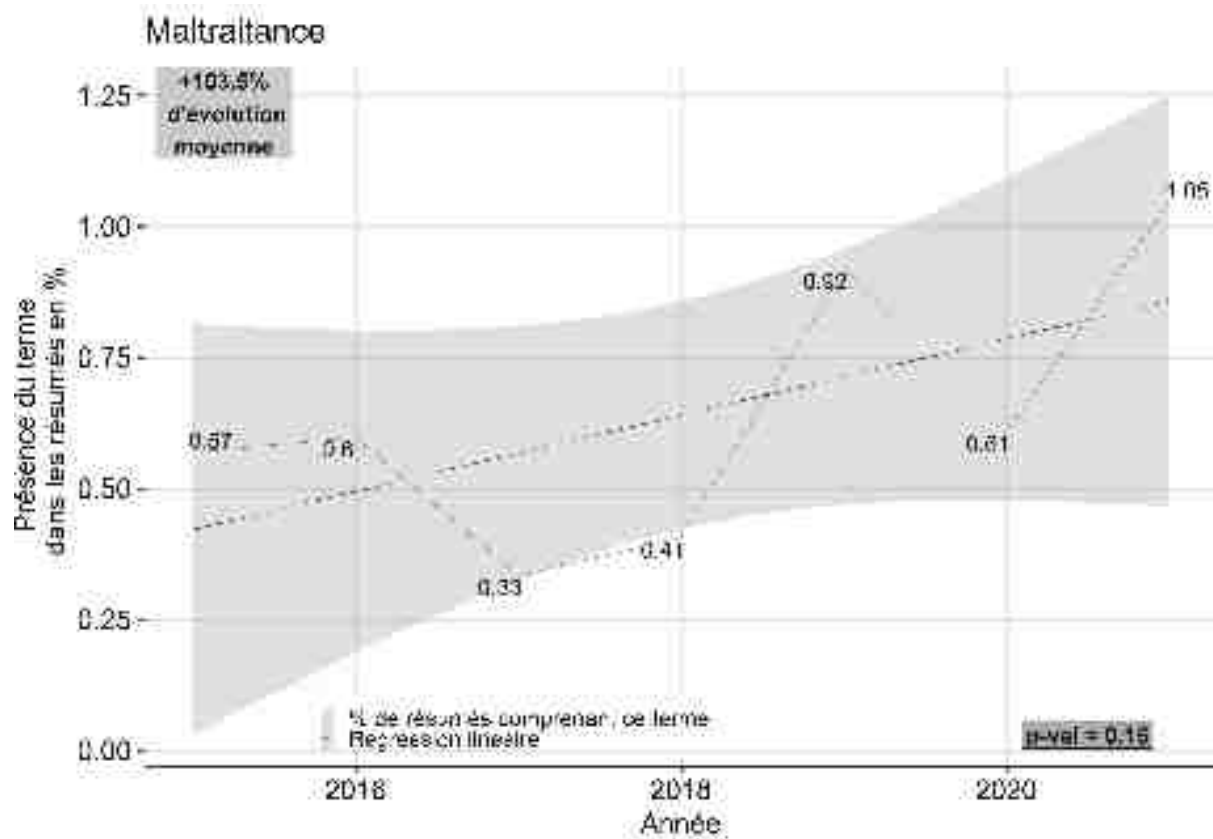
Je ne vois pas quelles précautions on peut recommander à ces Artisans, puisque la cause occasionnelle de leurs maux subsiste toujours, & qu'ils sont forcés de pourvoir à leurs besoins & à ceux de leur famille. Un purgatif, pris au Printemps & en Automne, peut empêcher qu'il ne s'amasse chez eux une si grande

(a) *Lumbi sedendo, oculi spectando dolent.*

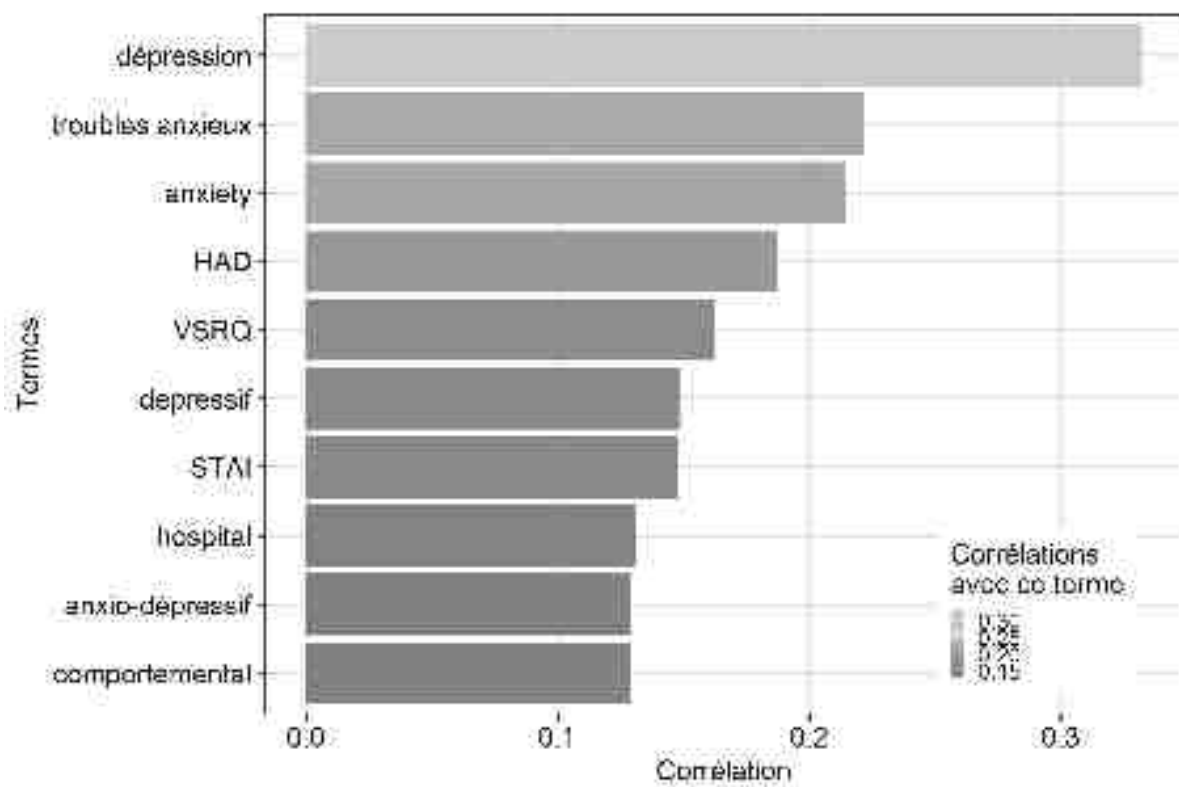
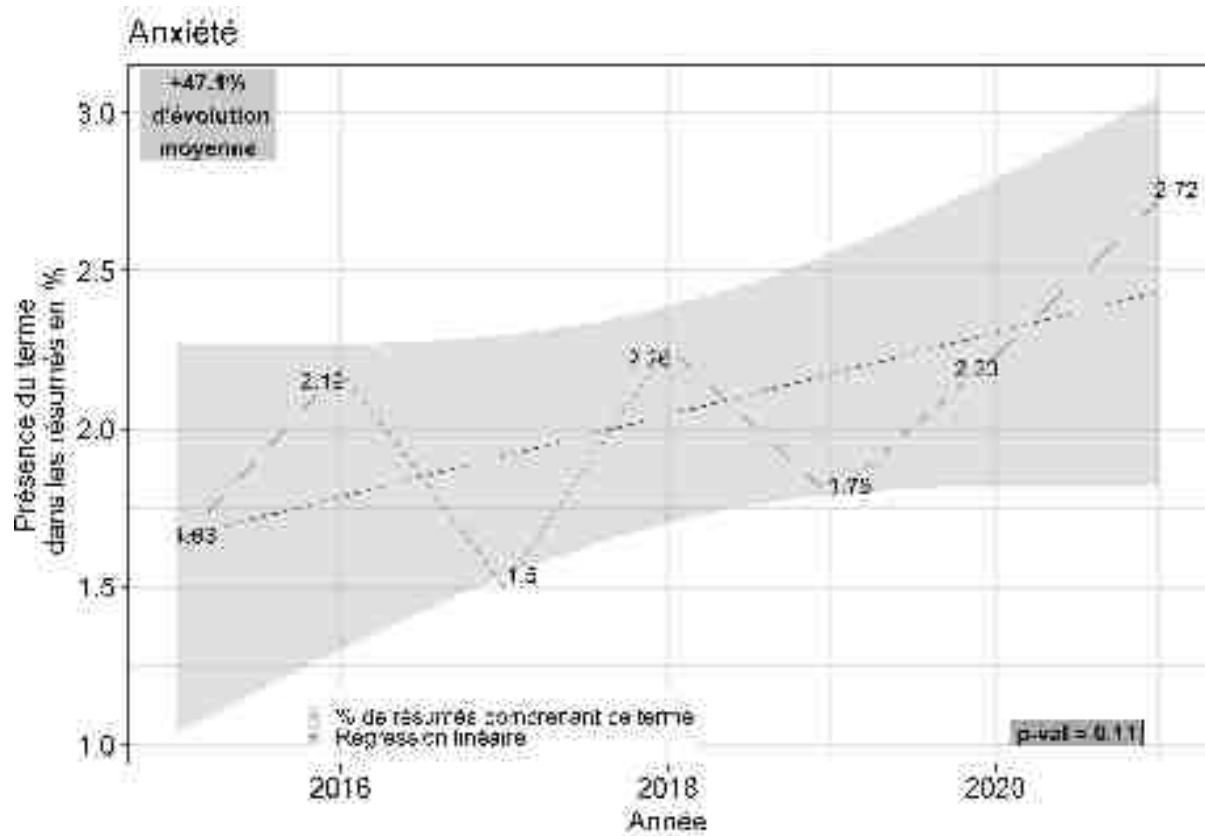
Annexe 6 : Évolution et corrélations de l'expression infections sexuellement transmissibles sur la période 2015 à 2021



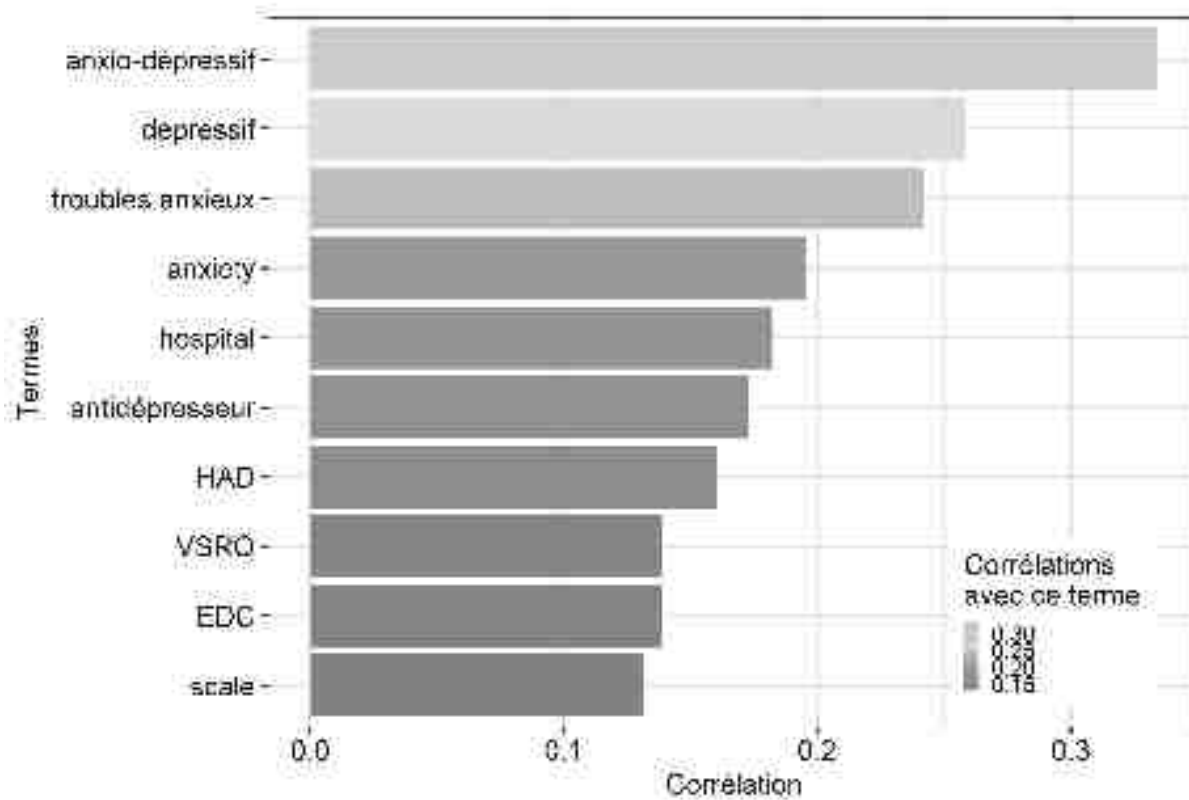
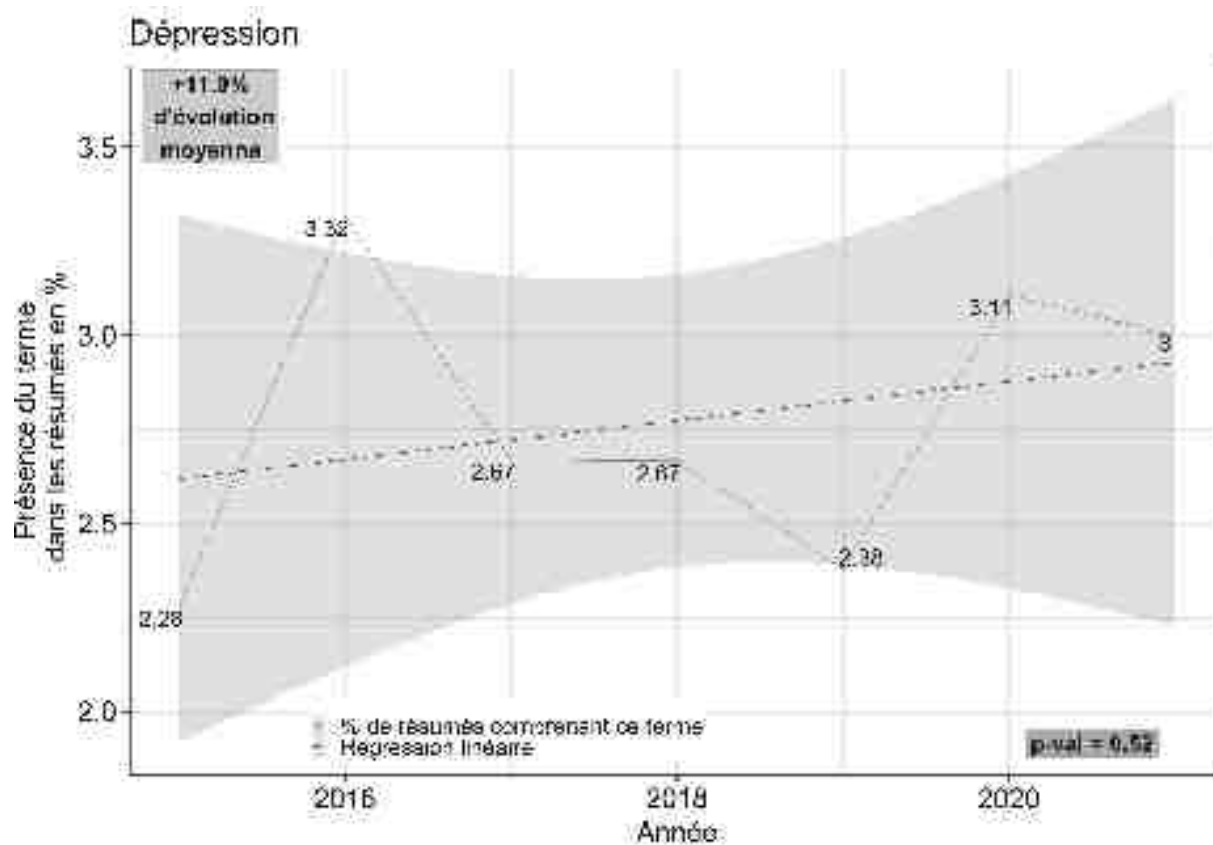
Annexe 7 : Évolution et corrélations du terme maltraitance sur la période 2015 à 2021



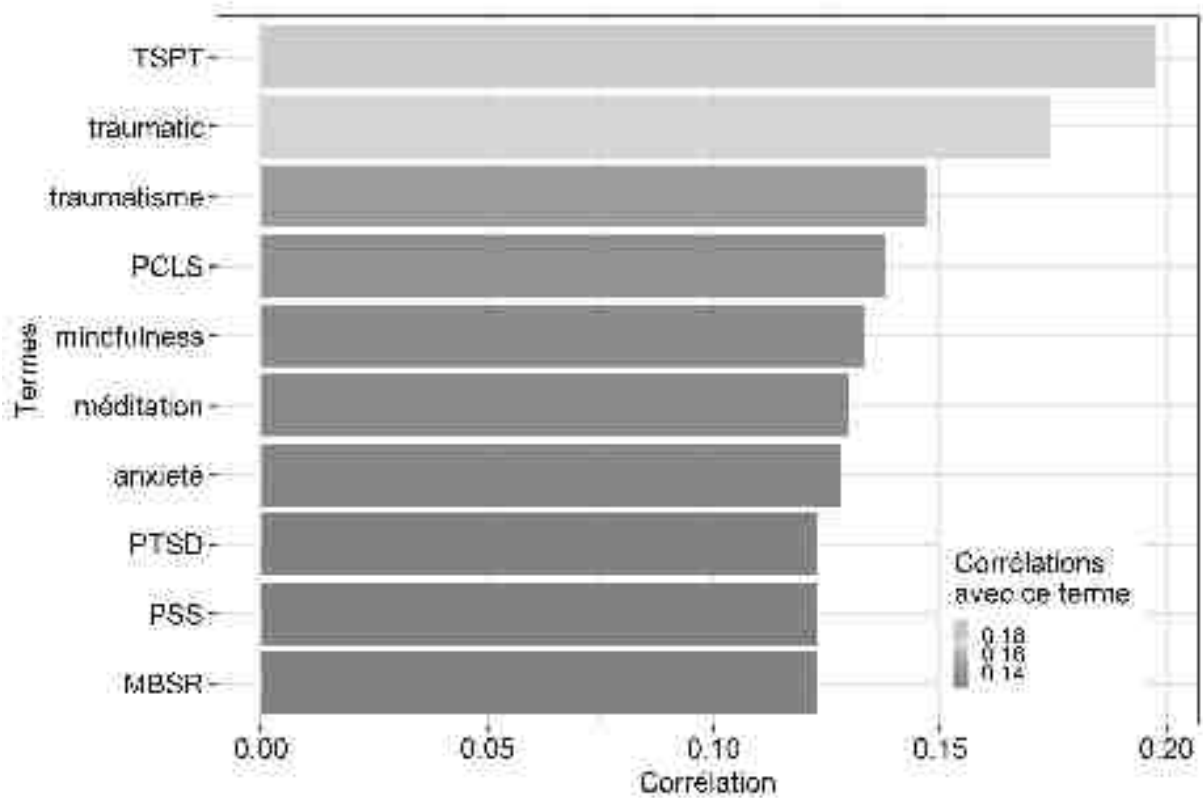
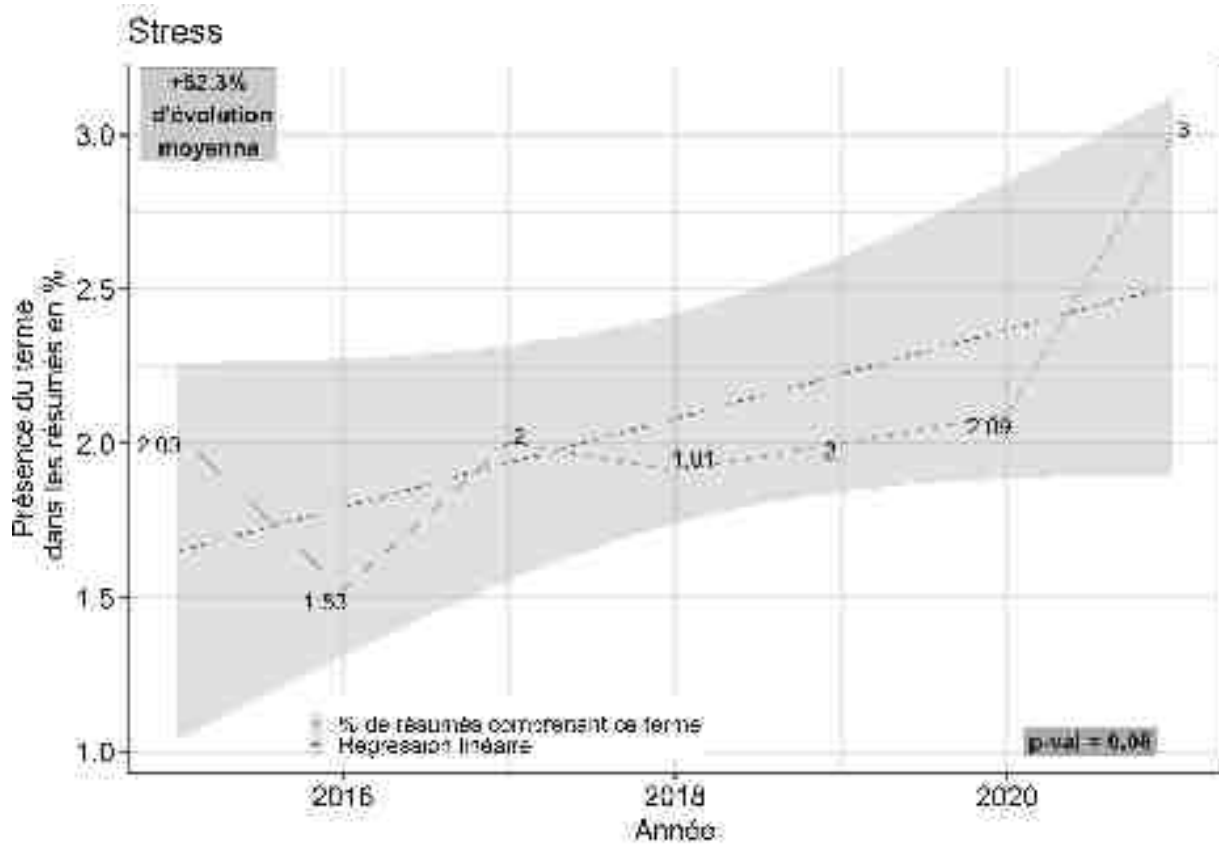
Annexe 8 : Évolution et corrélations du terme anxiété sur la période 2015 à 2021



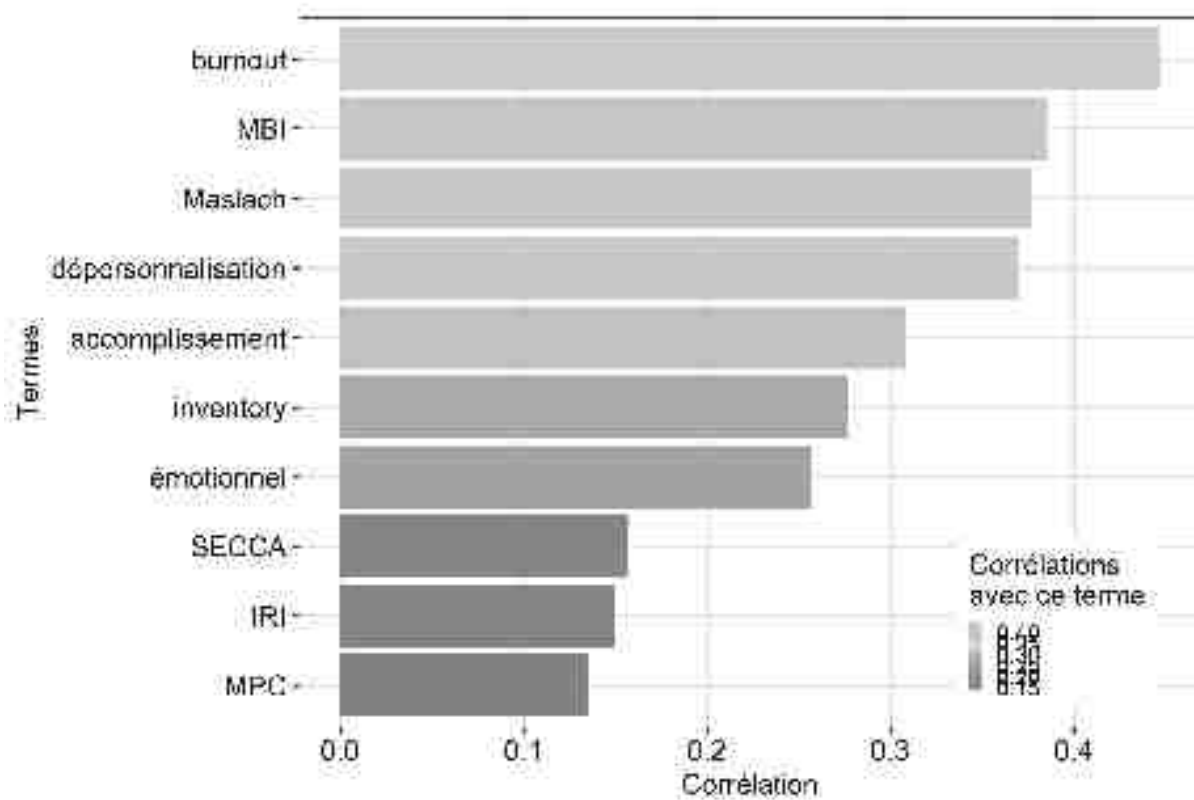
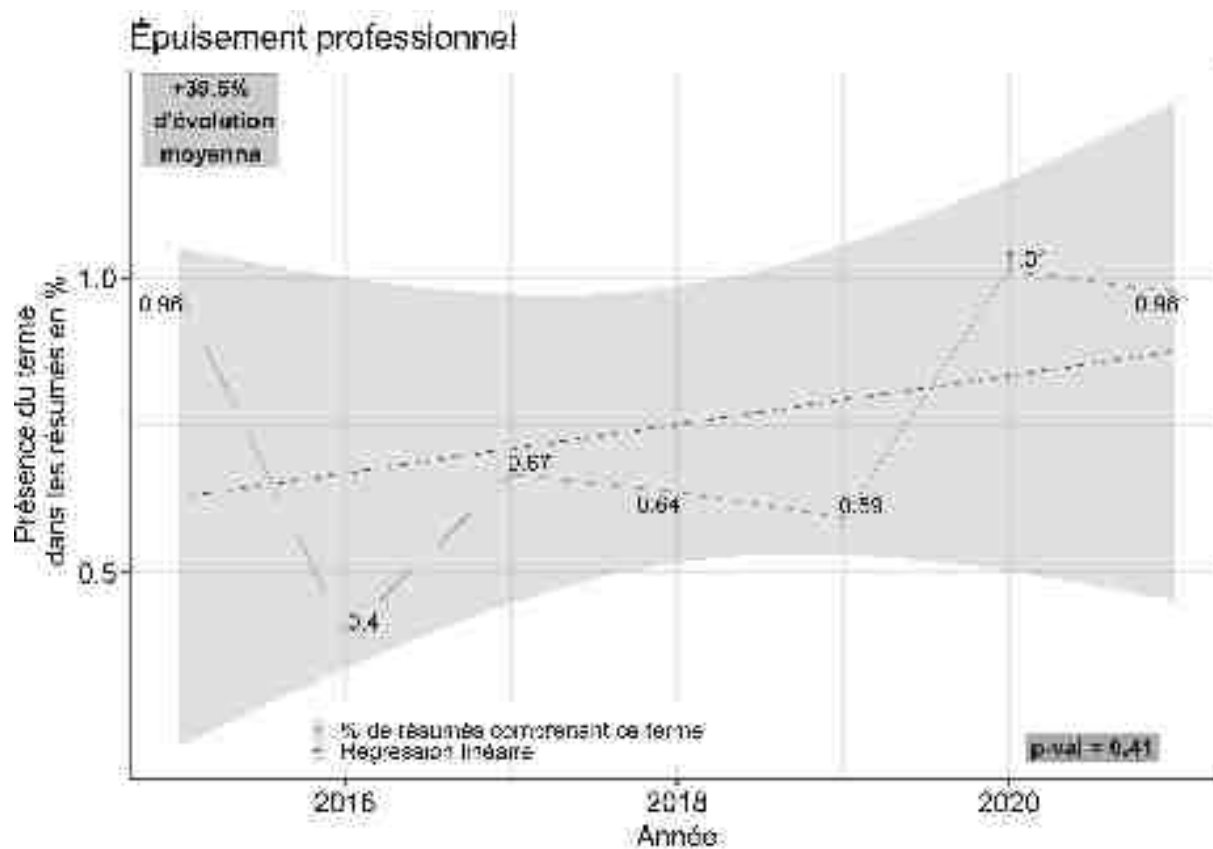
Annexe 9 : Évolution et corrélations du terme dépression sur la période 2015 à 2021



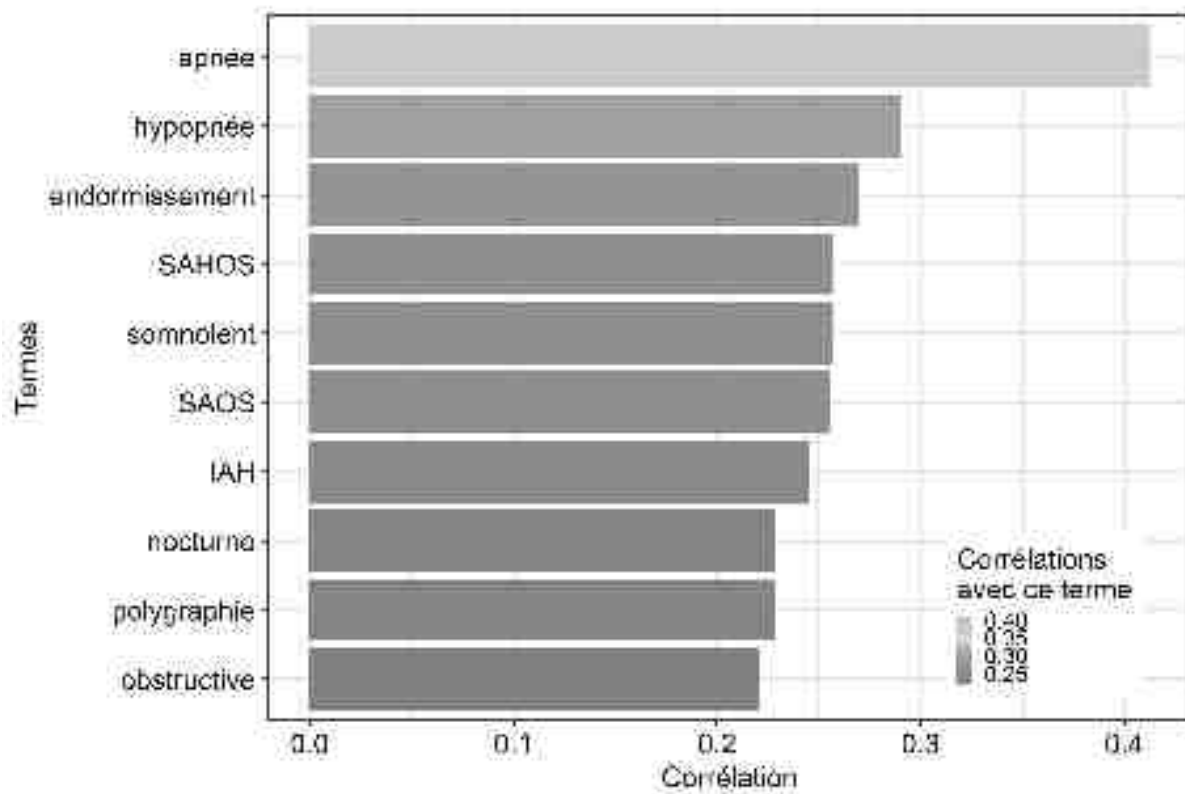
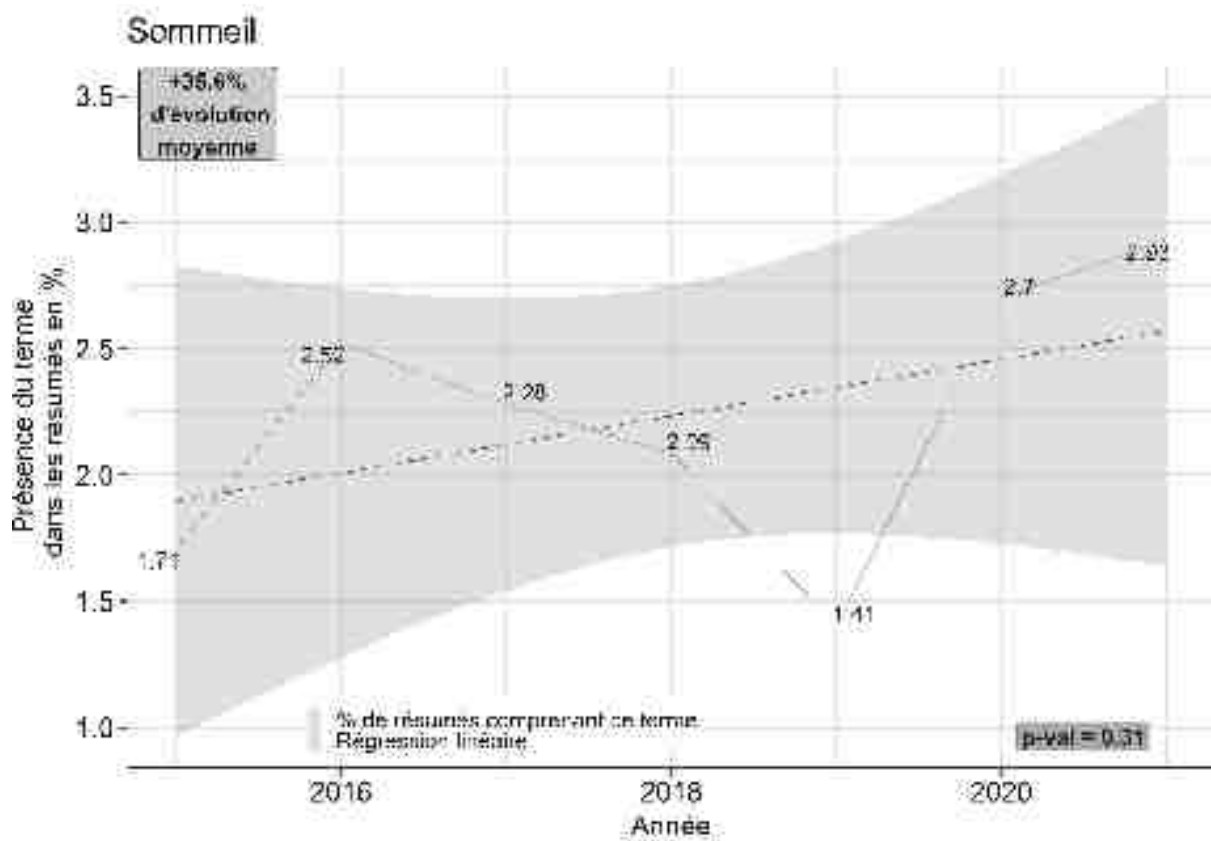
Annexe 10 : Évolution et corrélations du terme stress sur la période 2015 à 2021



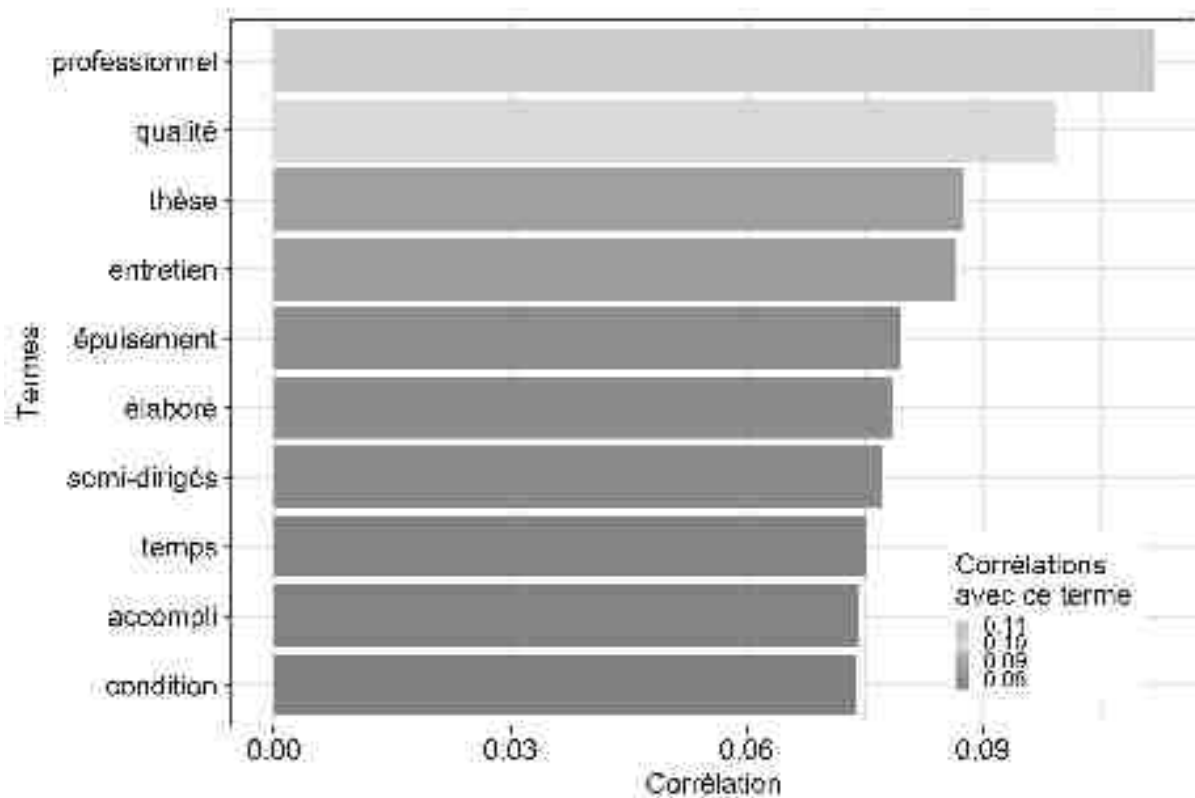
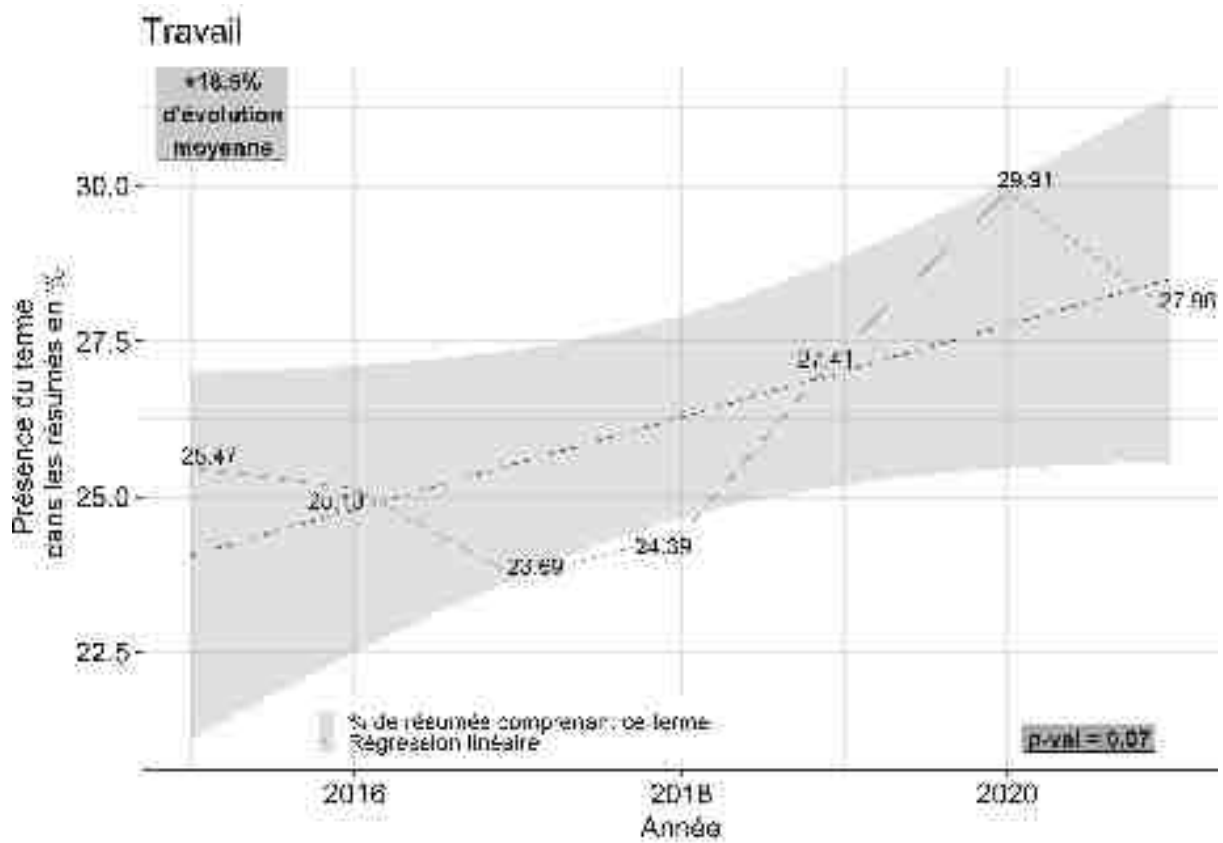
Annexe 11 : Évolution et corrélations de l'expression épuisement professionnel sur la période 2015 à 2021



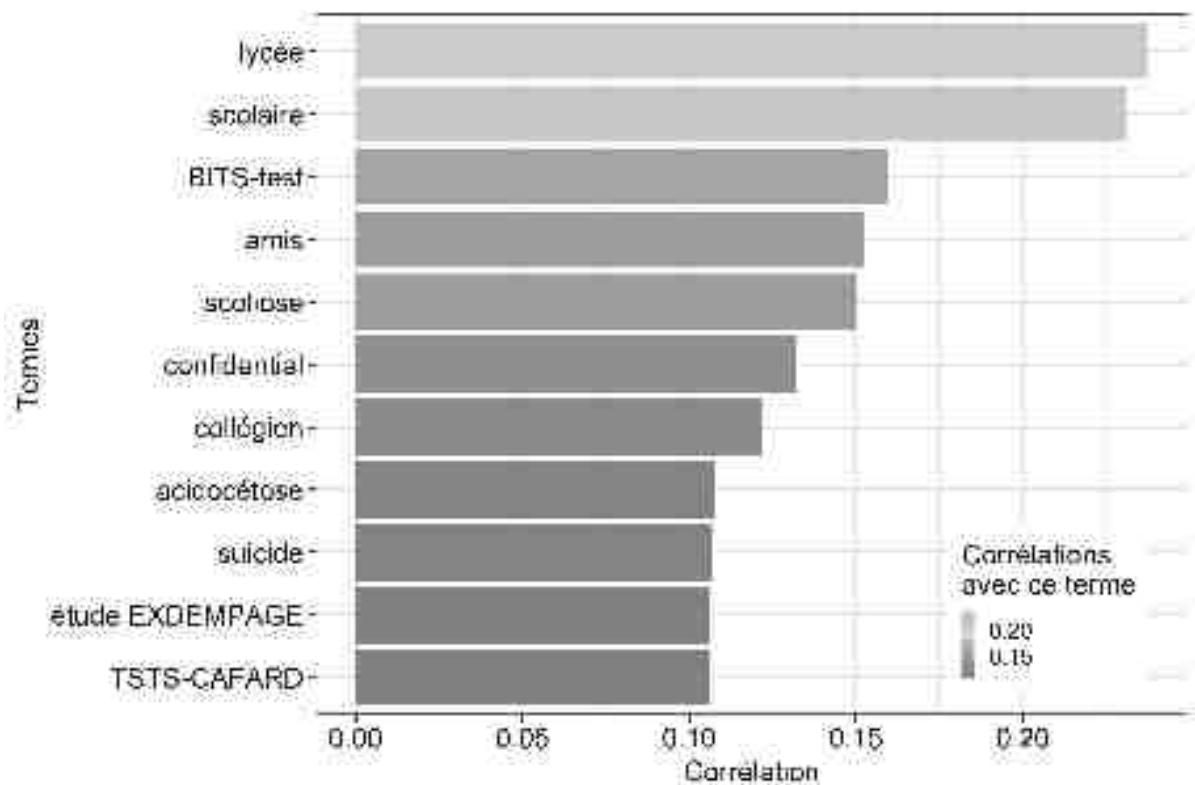
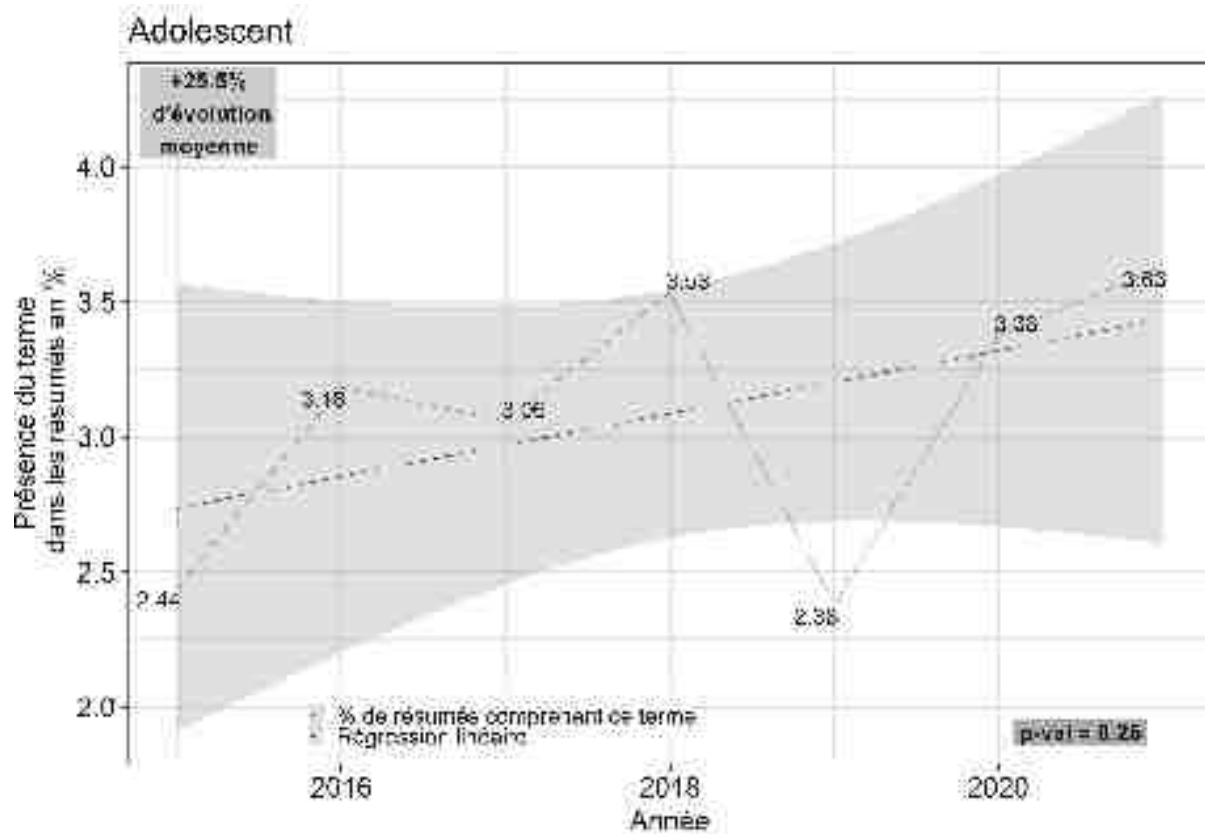
Annexe 12 : Évolution et corrélations du terme sommeil sur la période 2015 à 2021



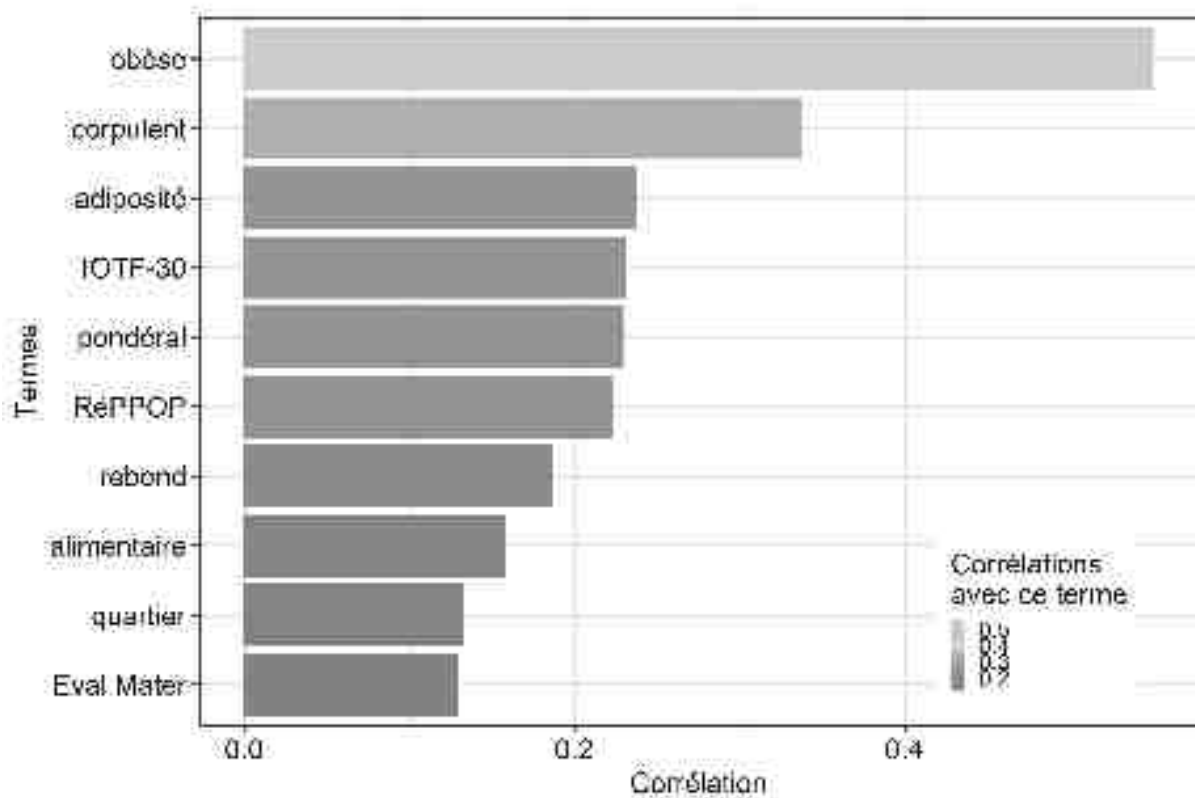
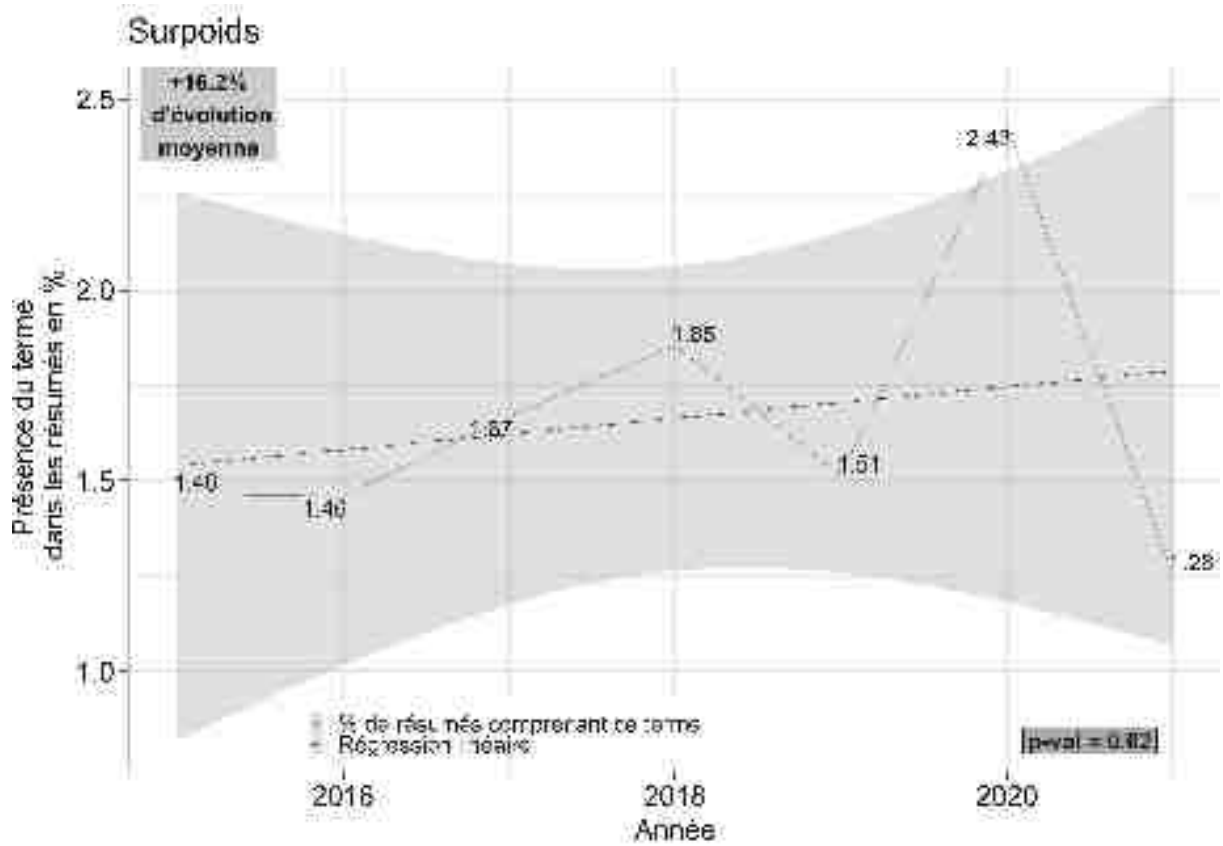
Annexe 13 : Évolution et corrélations du terme travail sur la période 2015 à 2021



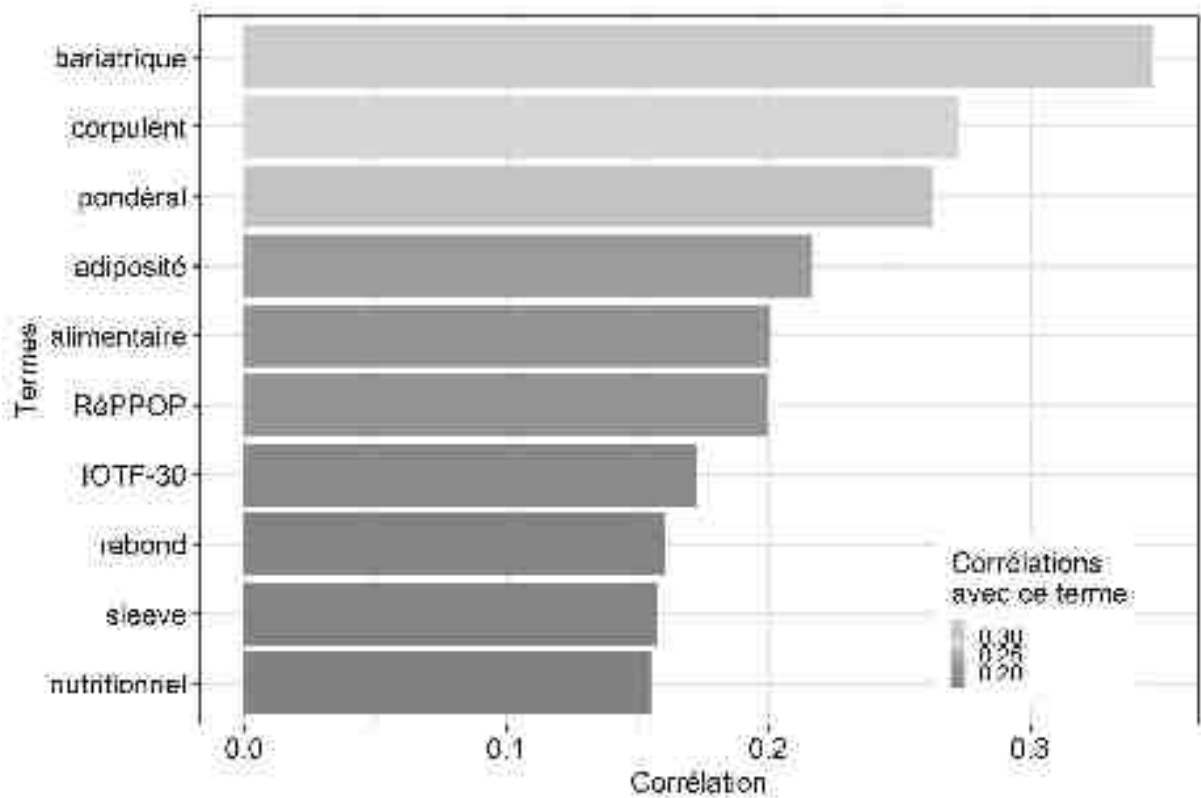
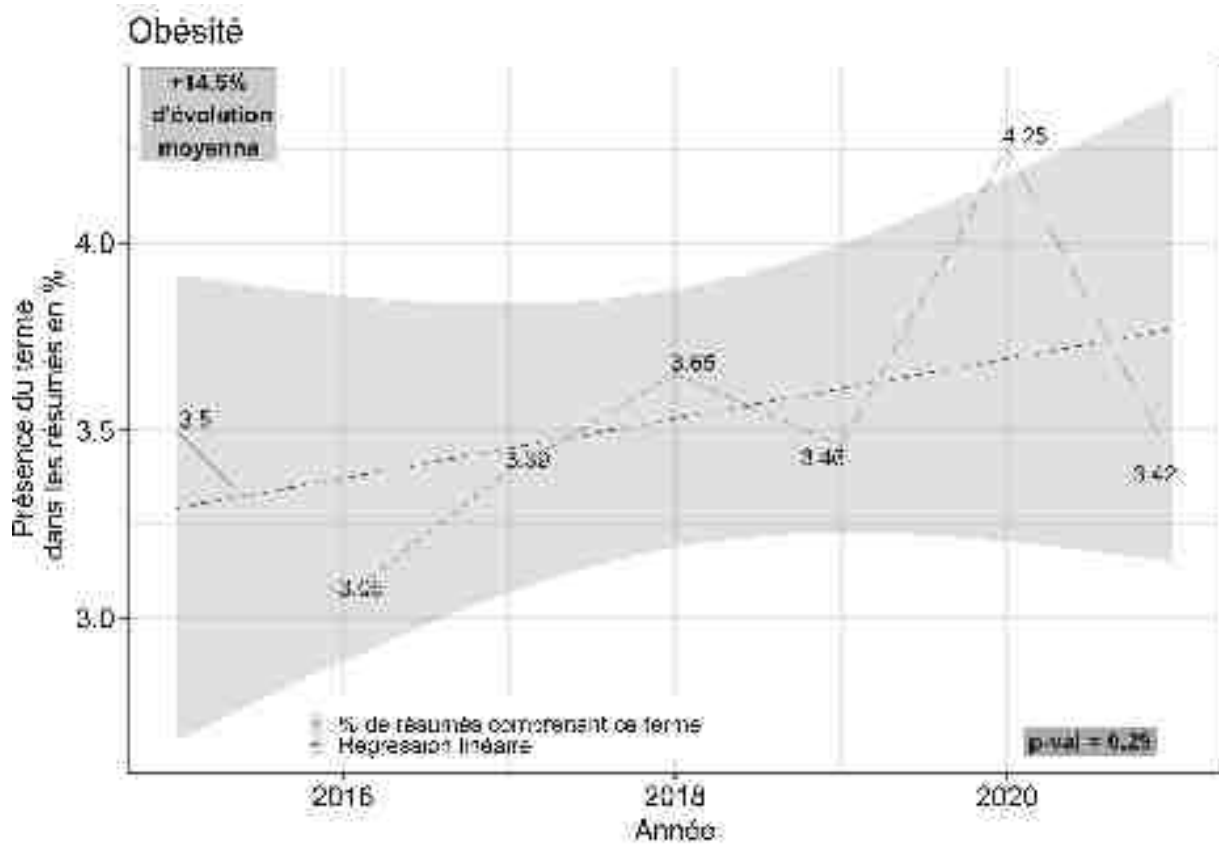
Annexe 14 : Évolution et corrélations du terme adolescent sur la période 2015 à 2021



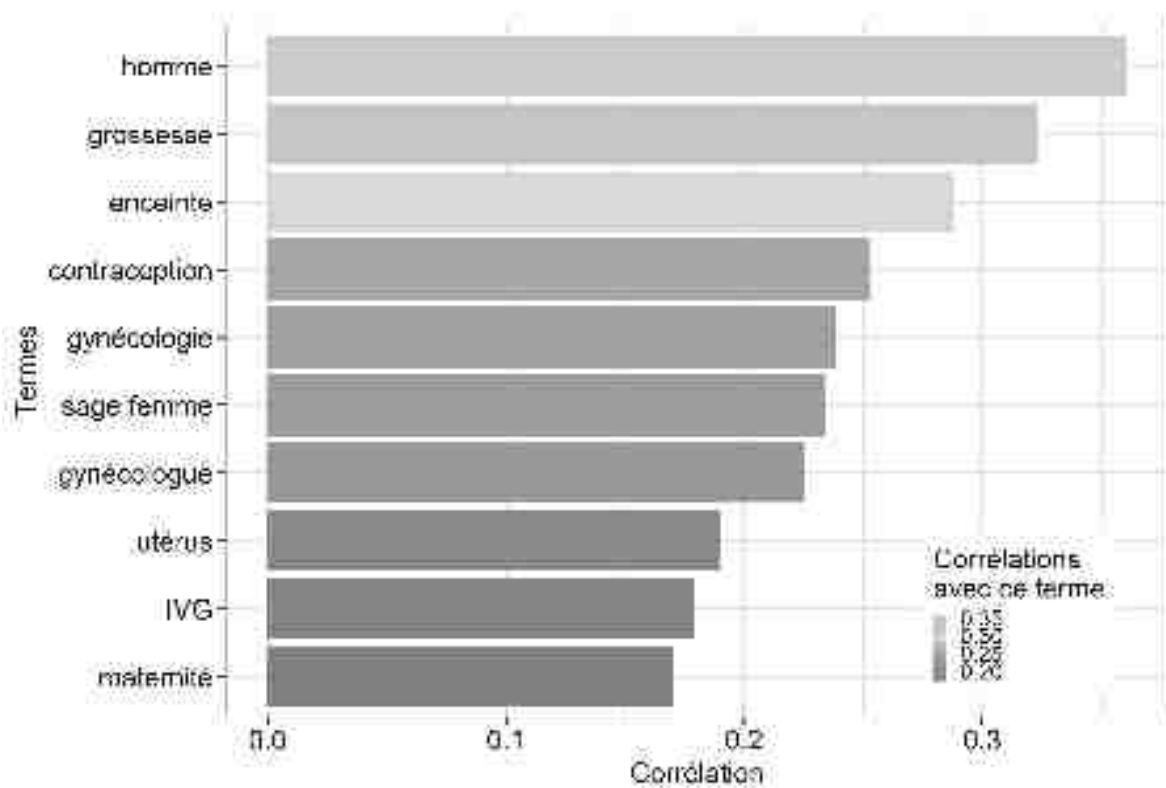
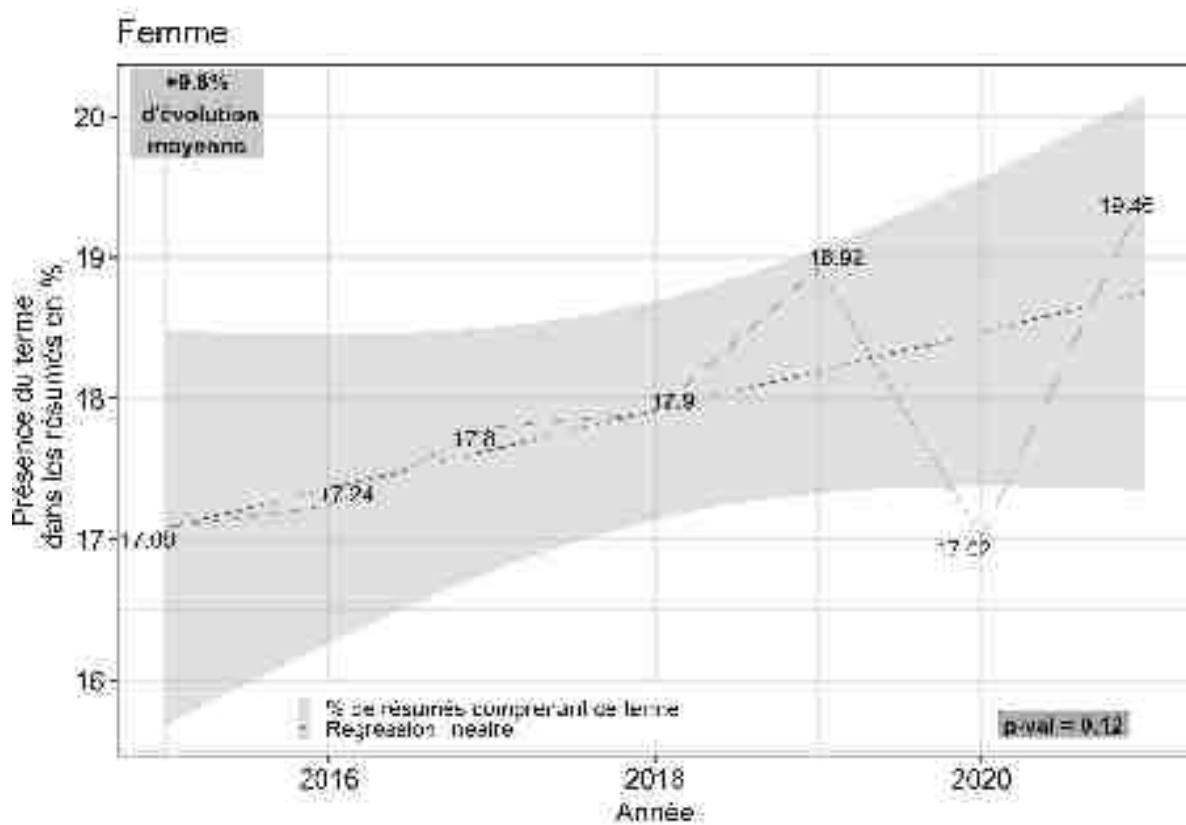
Annexe 15 : Évolution et corrélations du terme surpoids sur la période 2015 à 2021



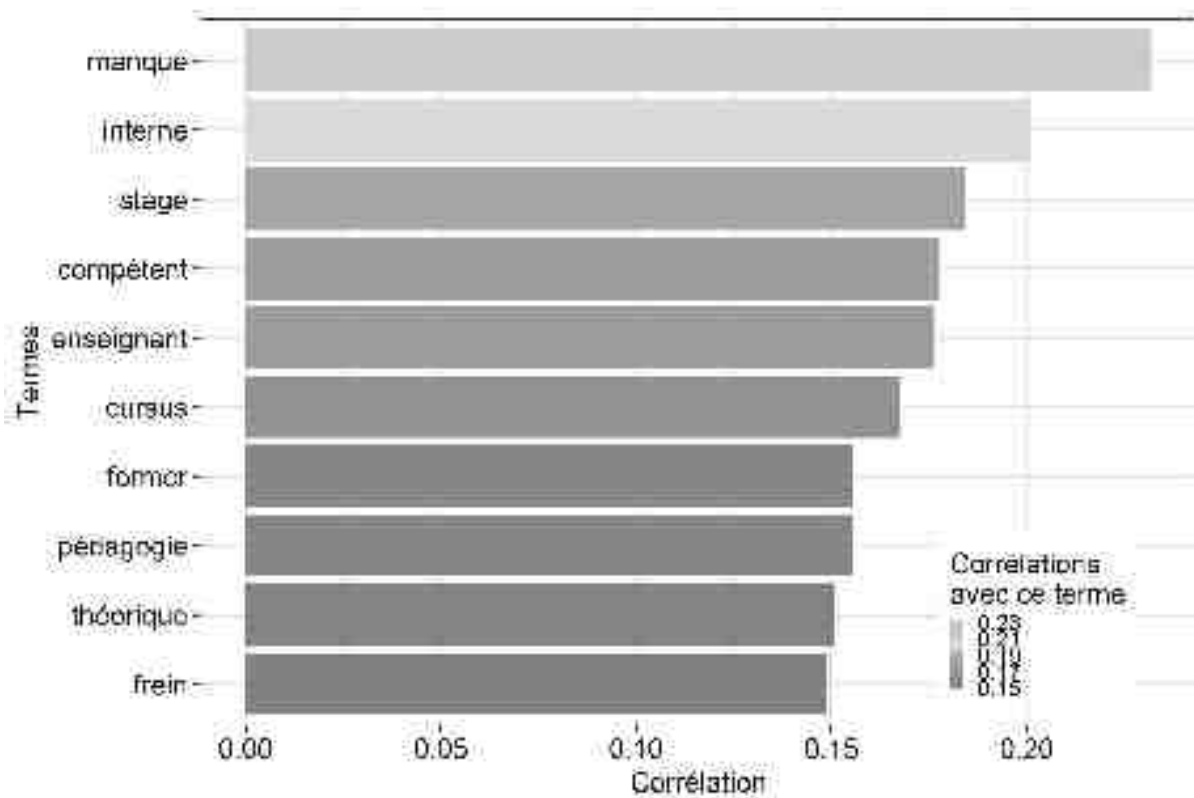
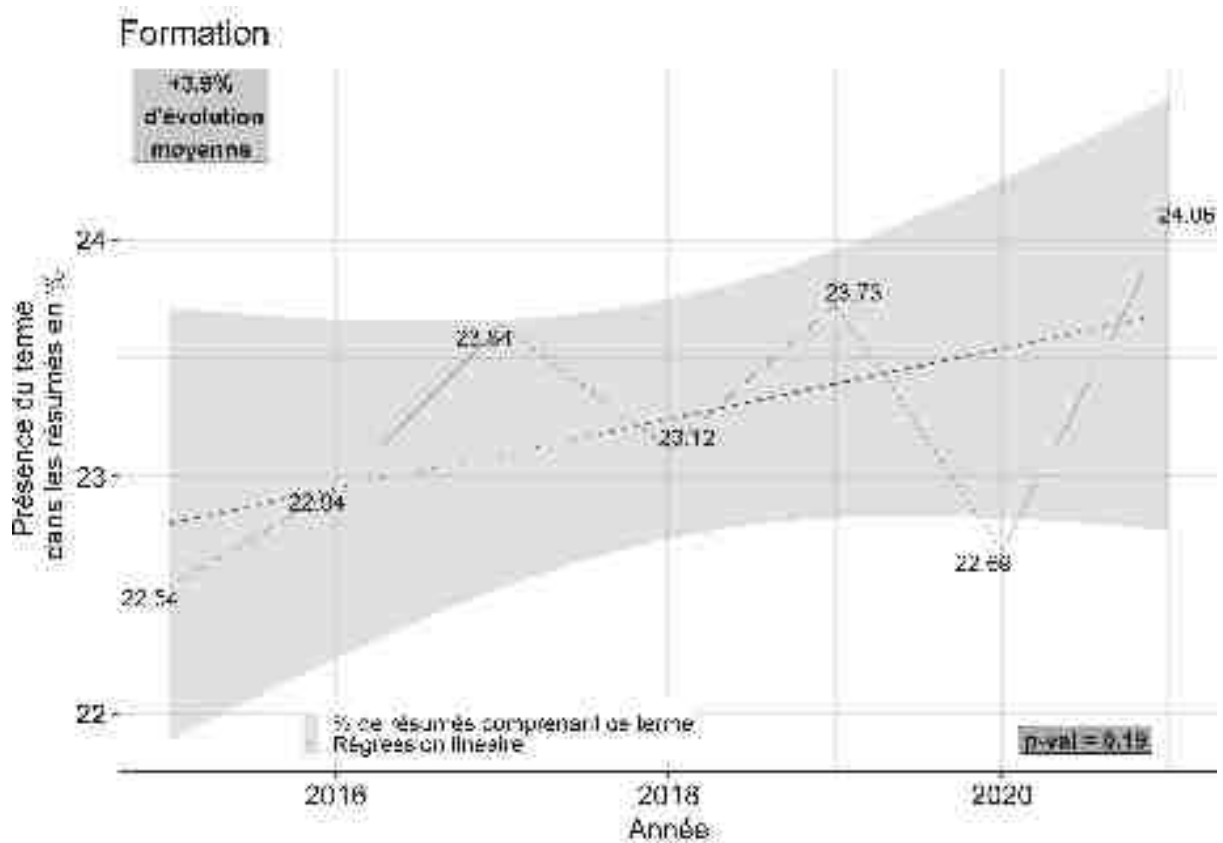
Annexe 16 : Évolution et corrélations du terme obésité sur la période 2015 à 2021



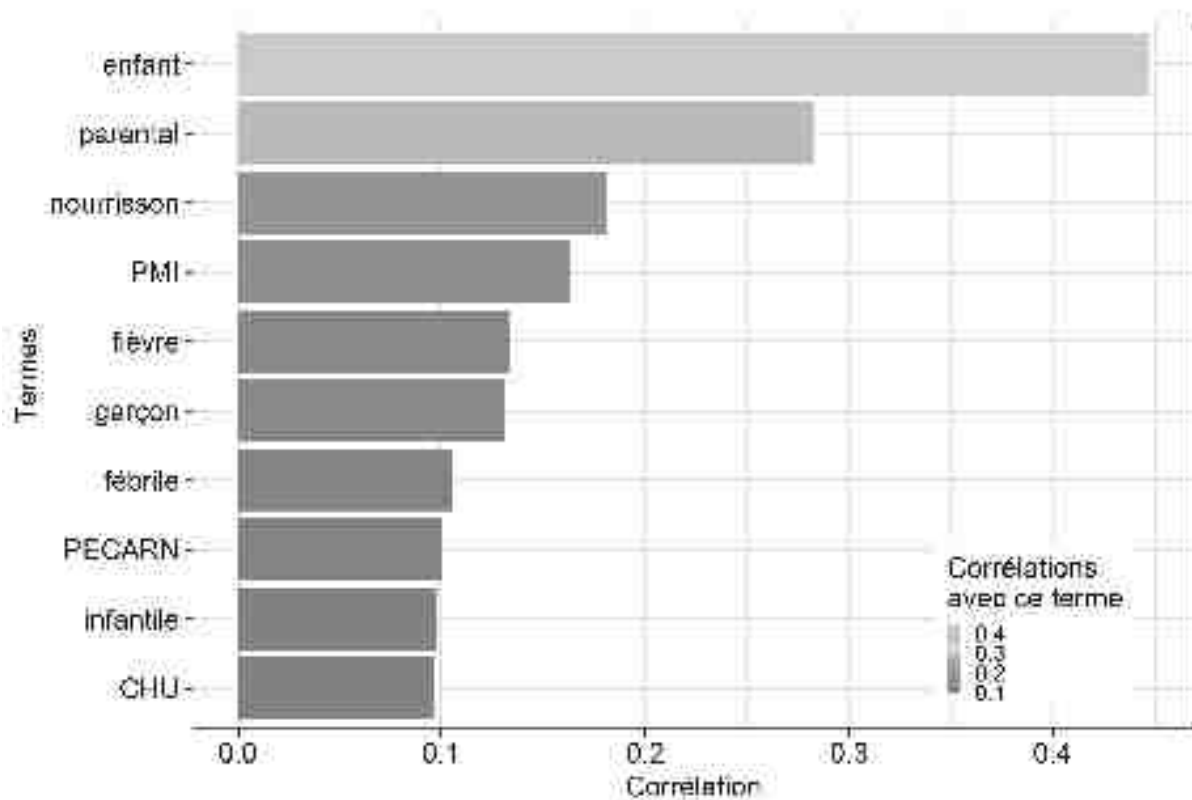
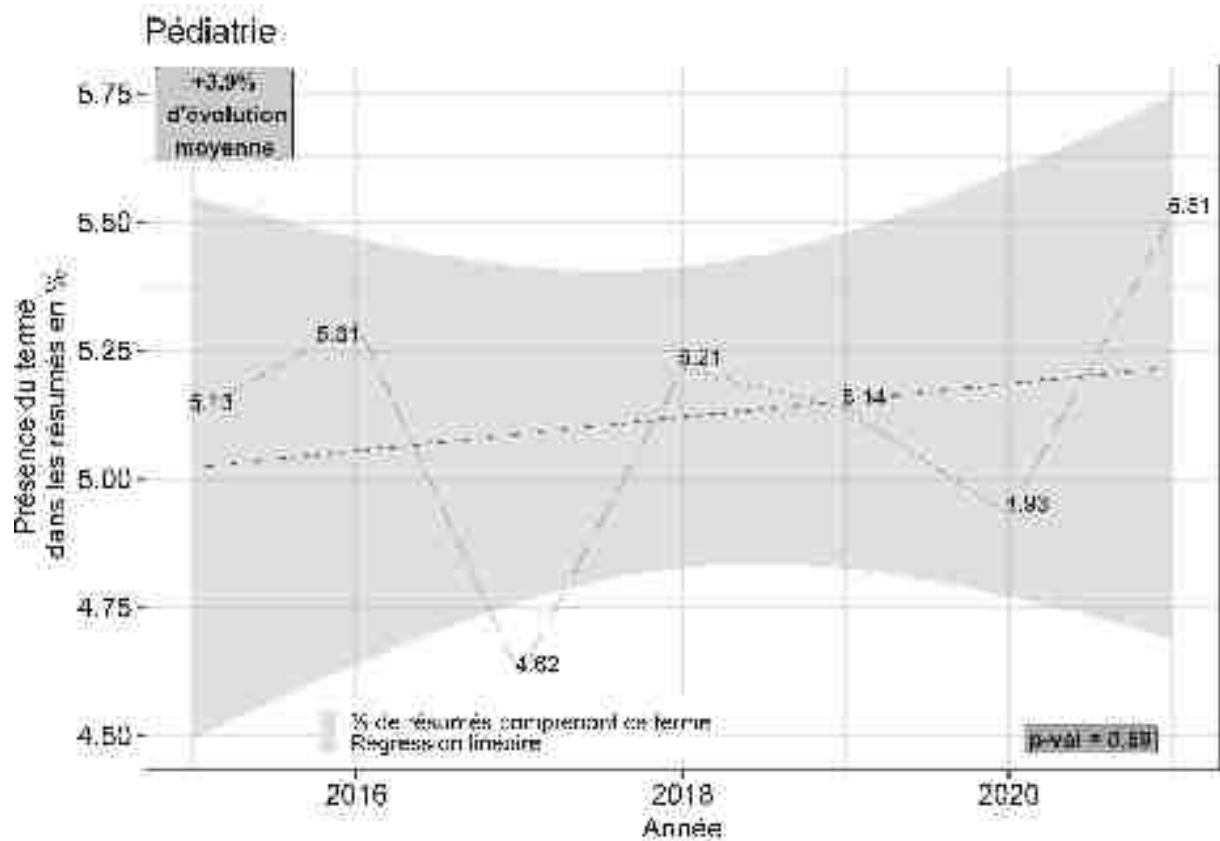
Annexe 17 : Évolution et corrélations du terme femme sur la période 2015 à 2021



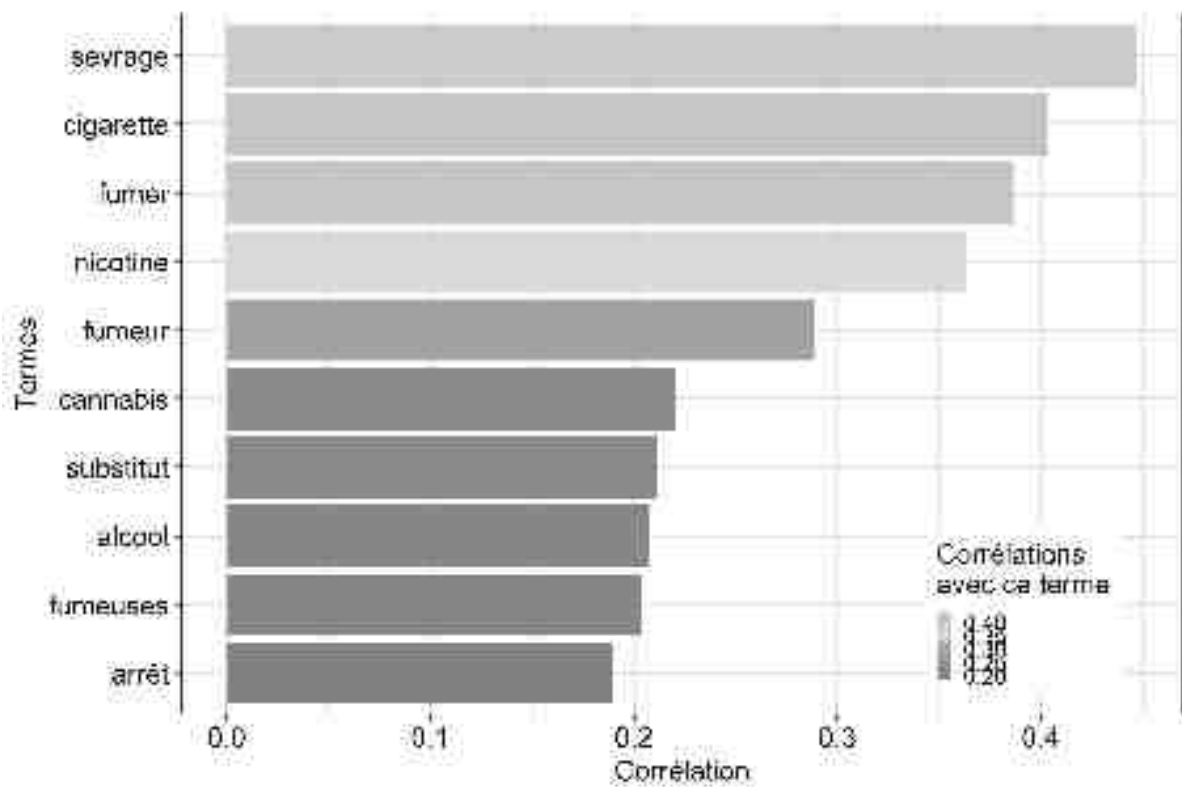
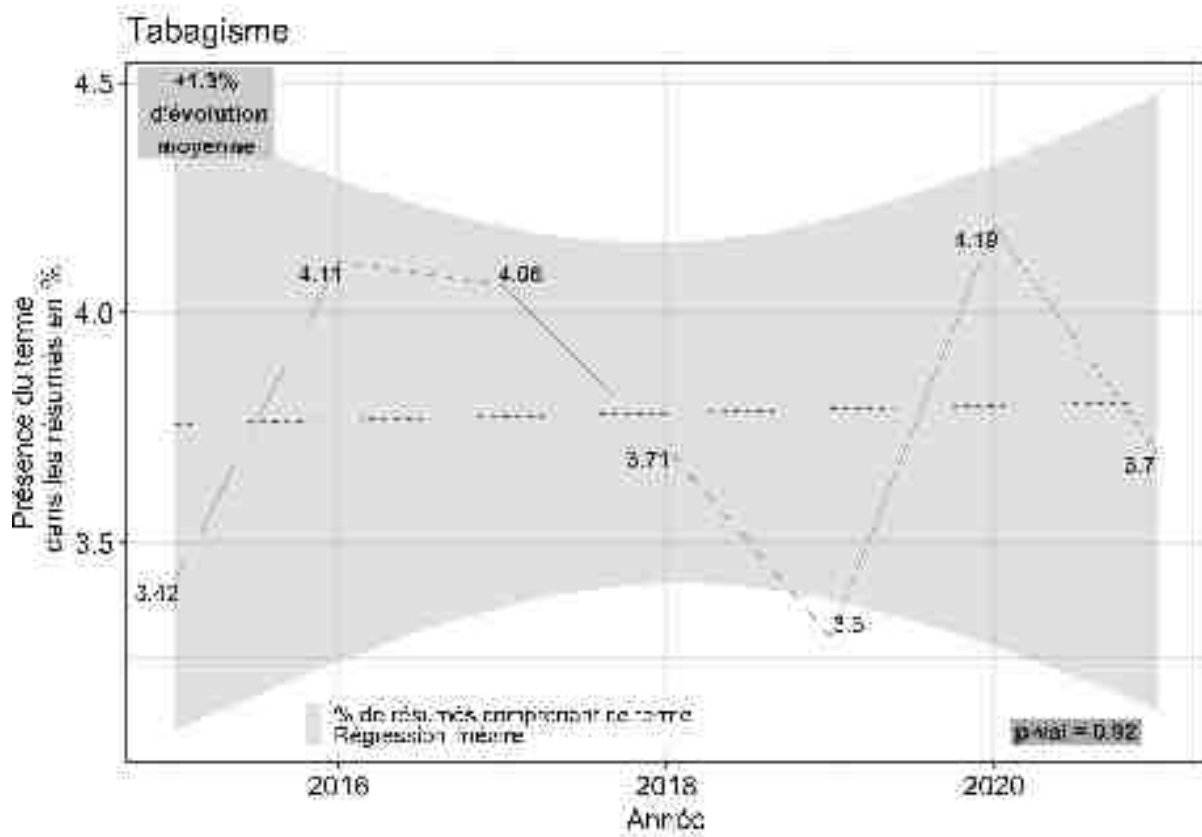
Annexe 18: Évolution et corrélations du terme formation sur la période 2015 à 2021



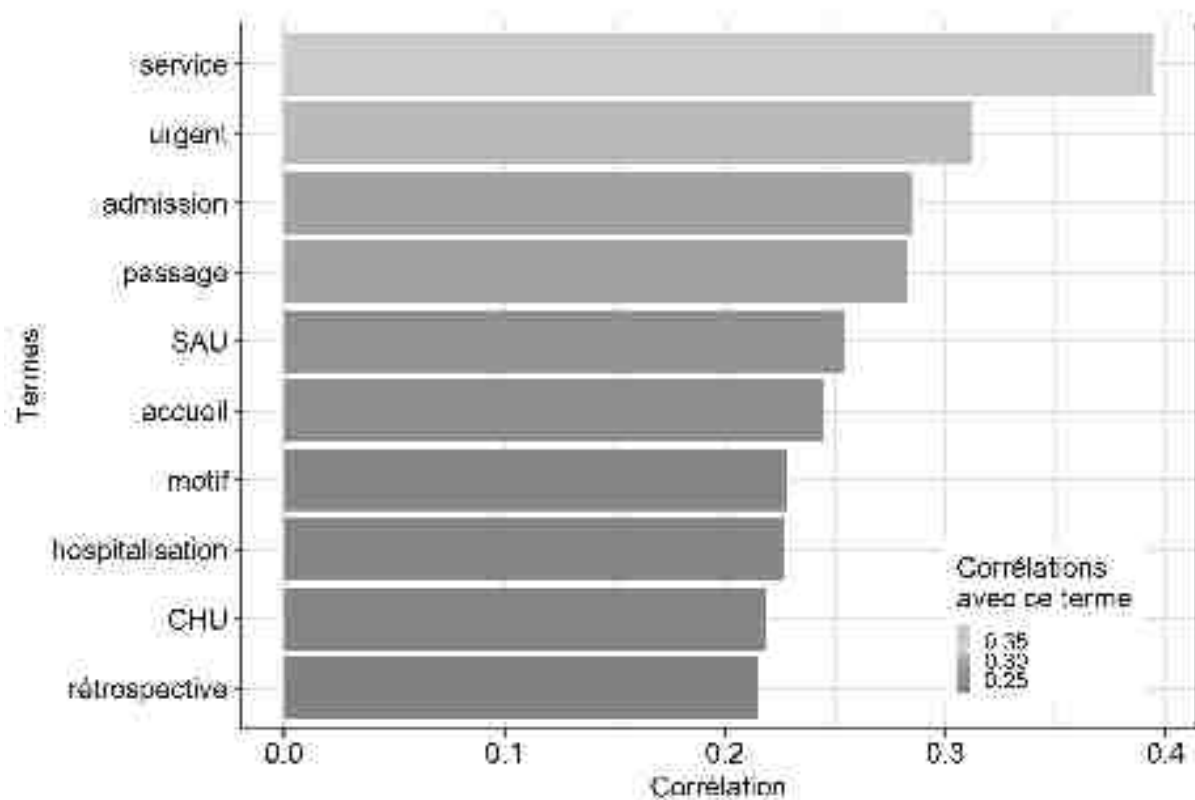
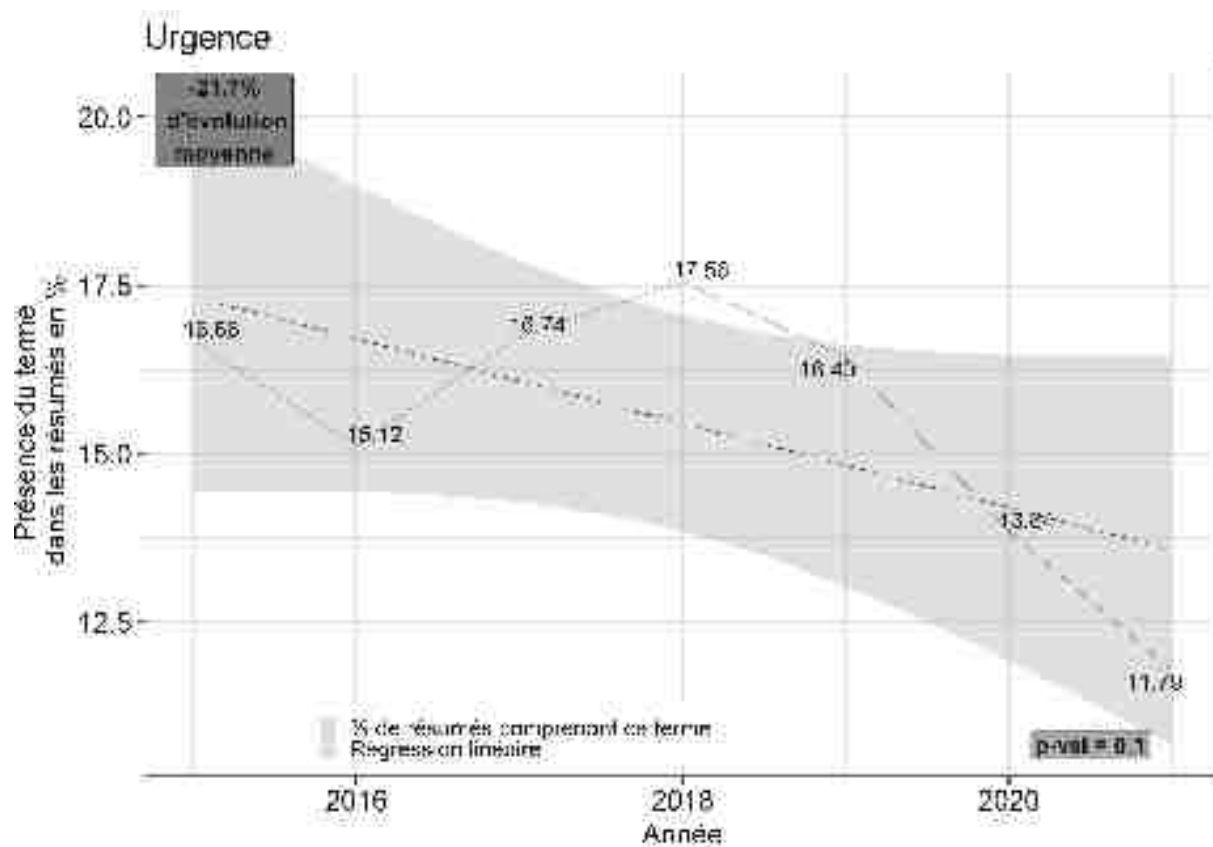
Annexe 19 : Évolution et corrélations du terme pédiatrie sur la période 2015 à 2021



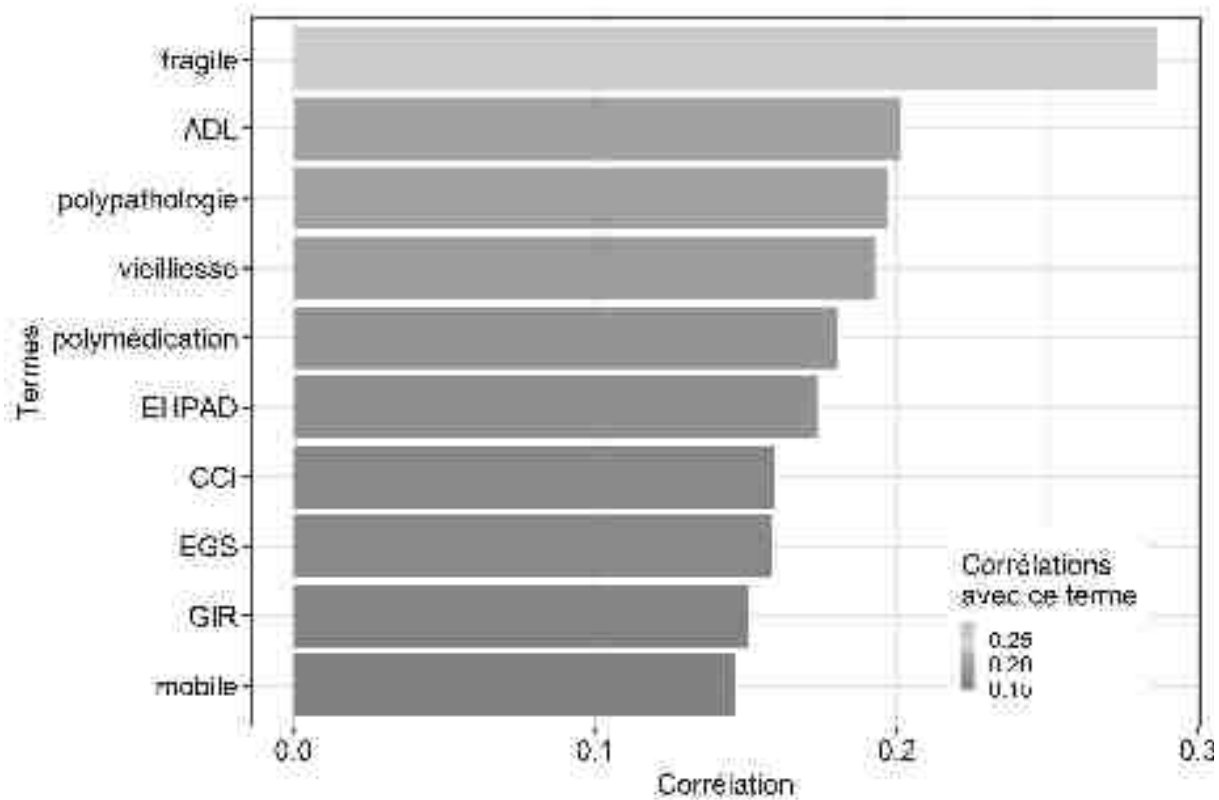
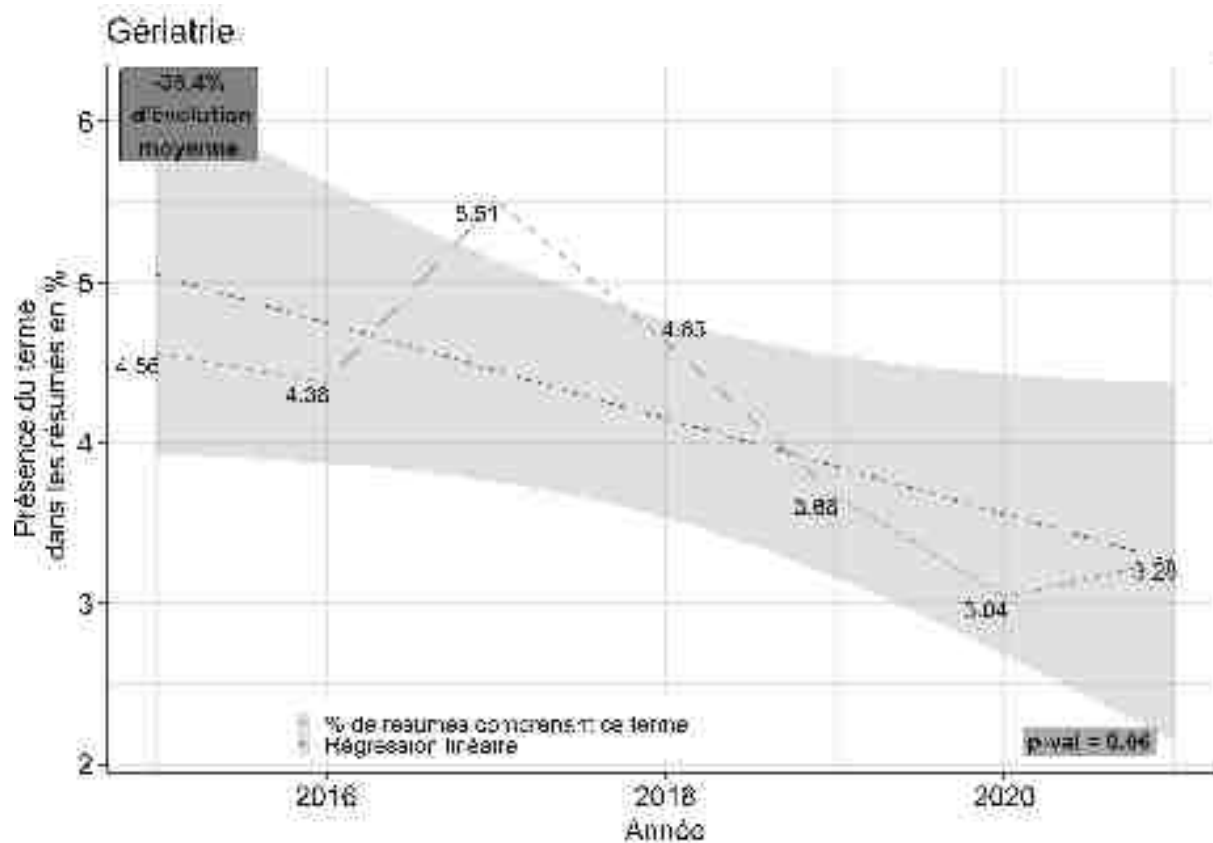
Annexe 20 : Évolution et corrélations du terme tabagisme sur la période 2015 à 2021



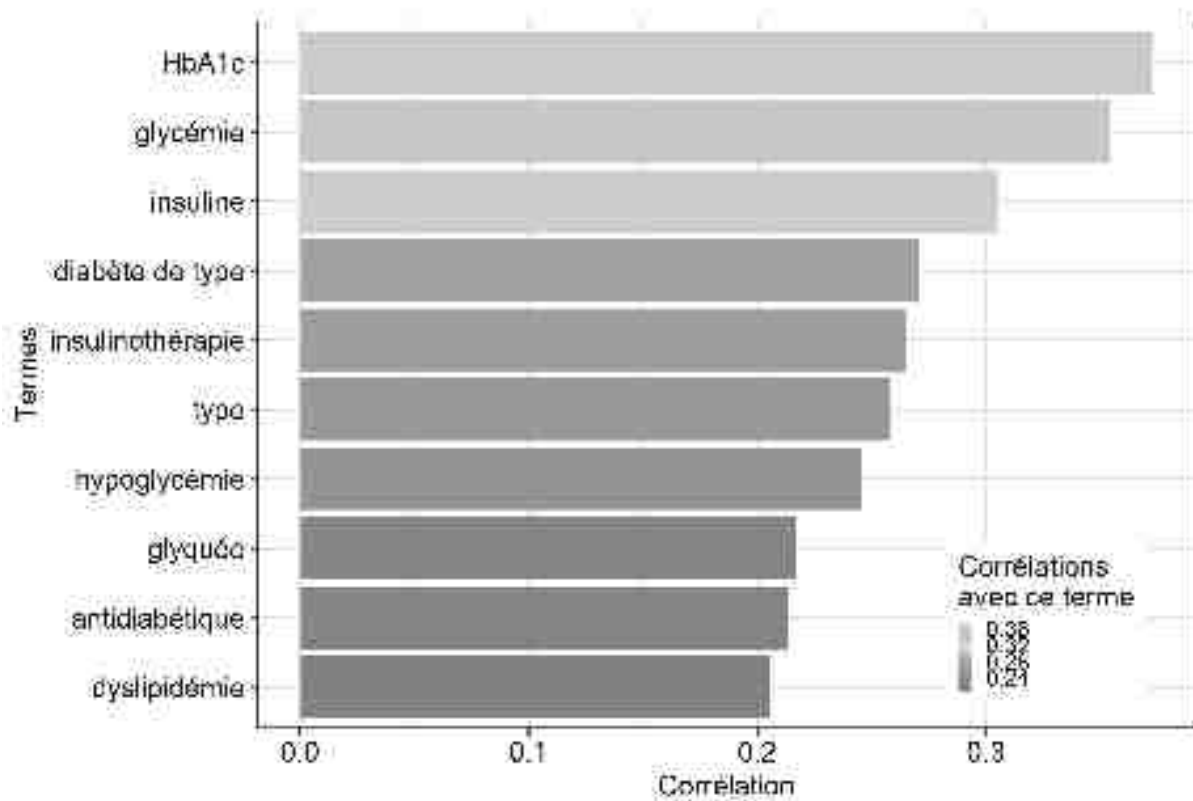
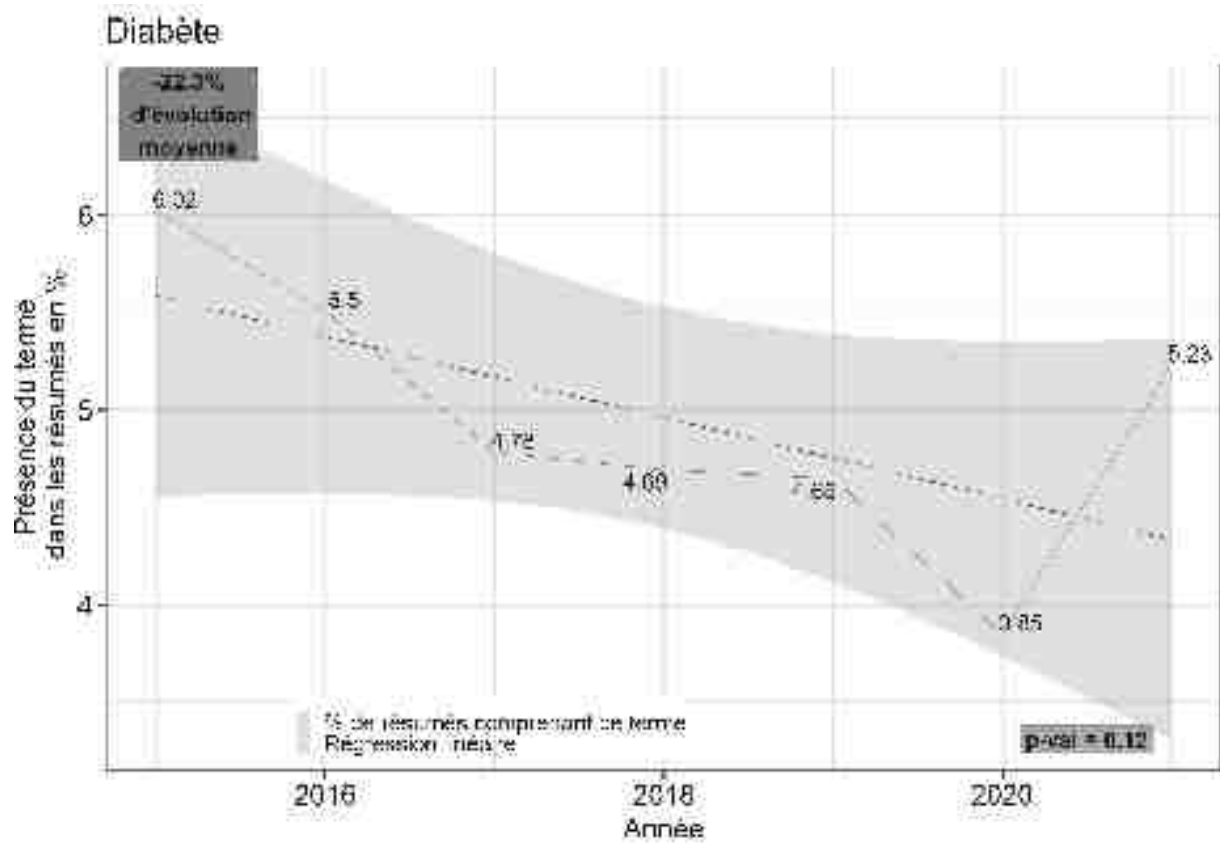
Annexe 21 : Évolution et corrélations du terme urgence sur la période 2015 à 2021



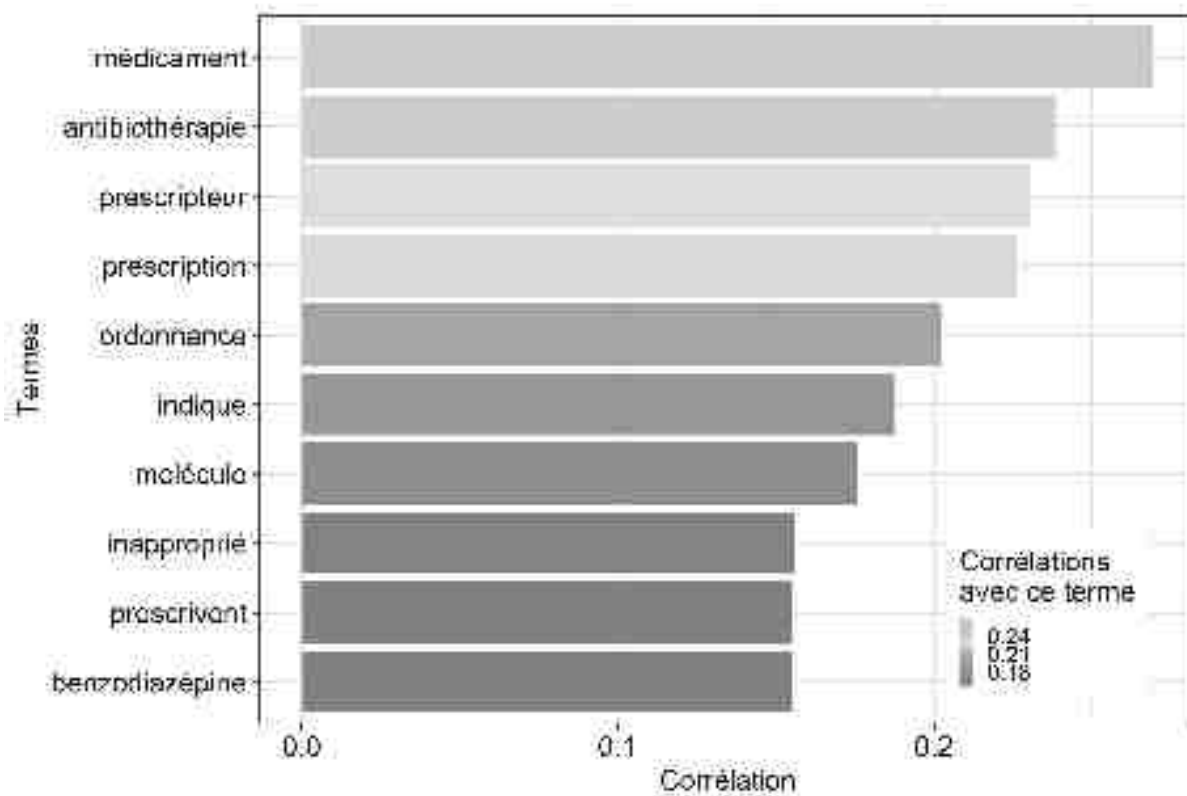
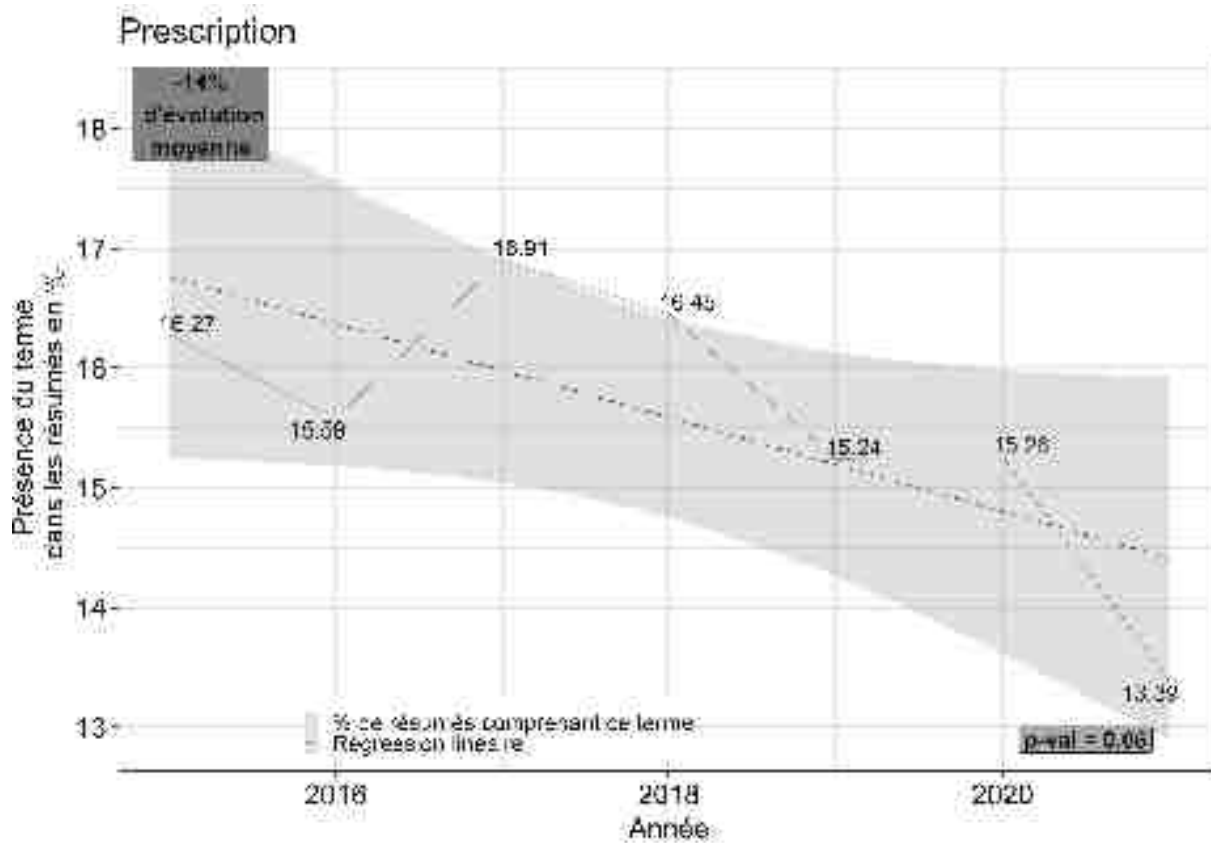
Annexe 22 : Évolution et corrélations du terme gériatrie sur la période 2015 à 2021



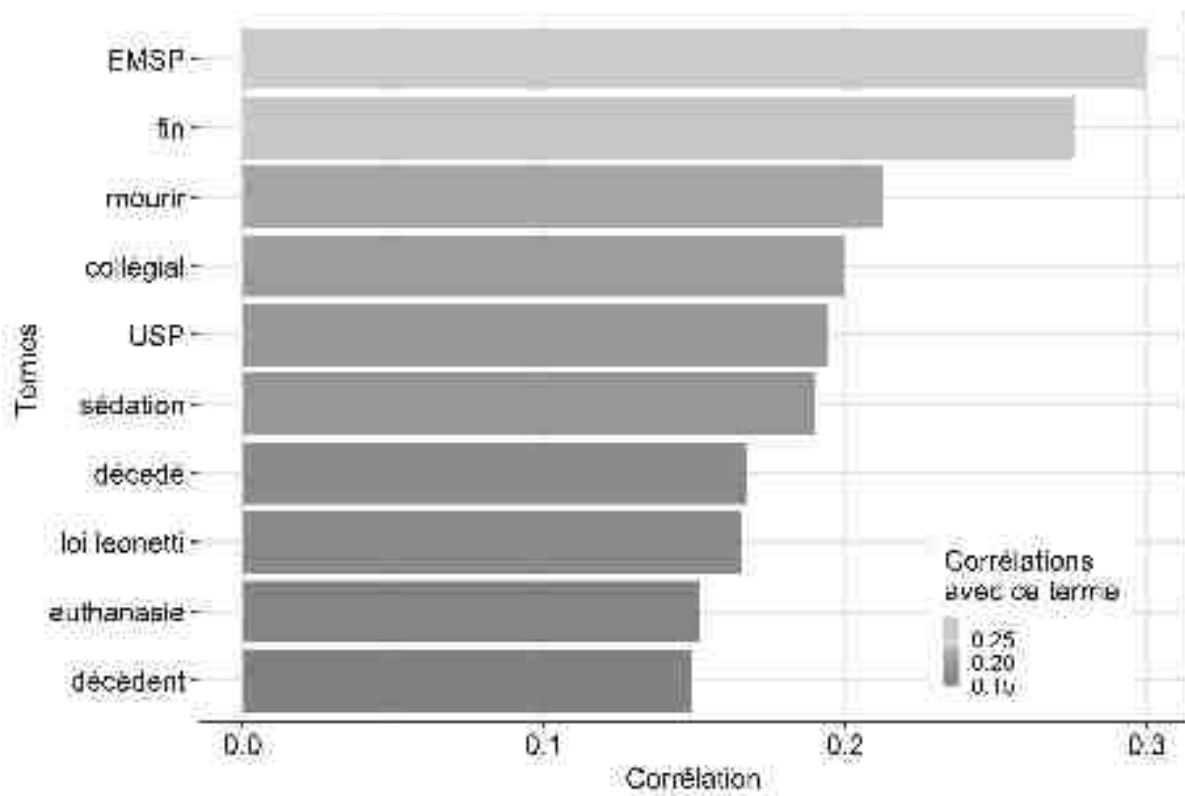
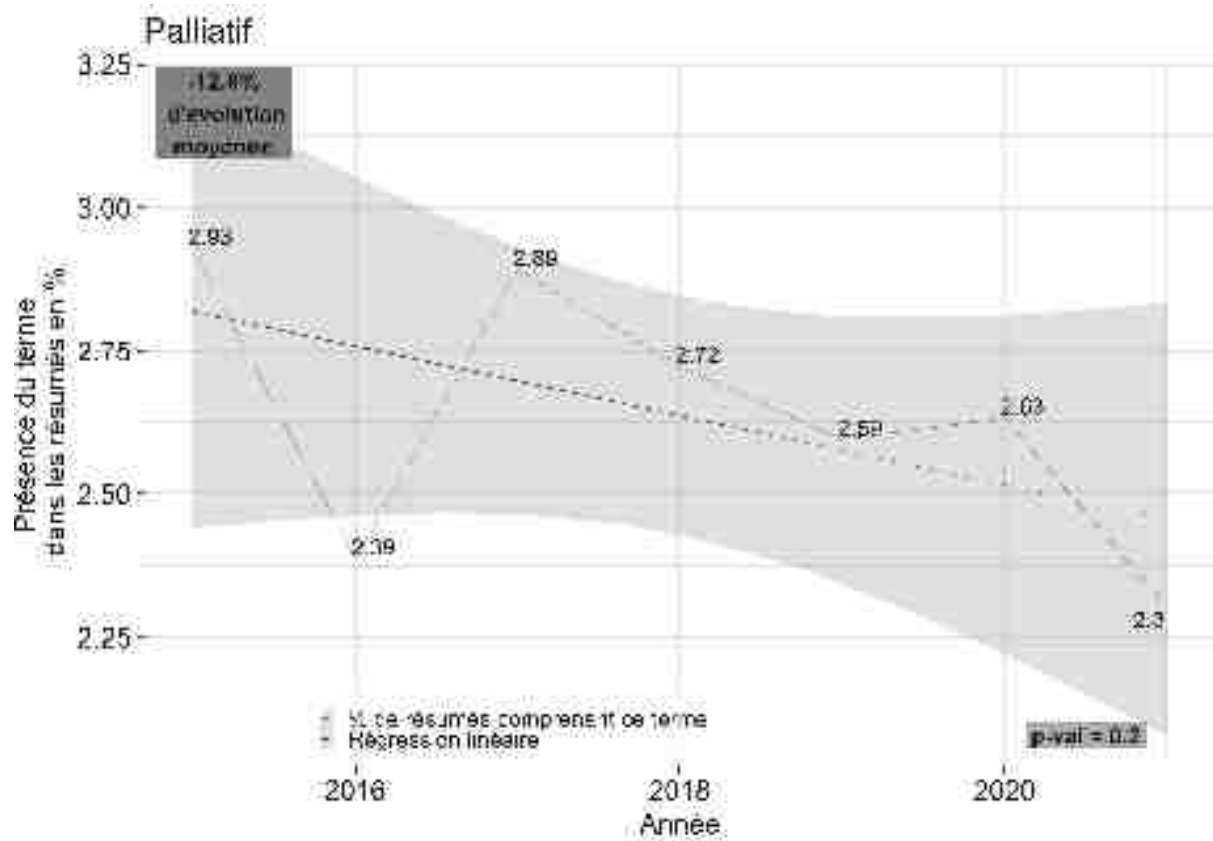
Annexe 23 : Évolution et corrélations du terme diabète sur la période 2015 à 2021



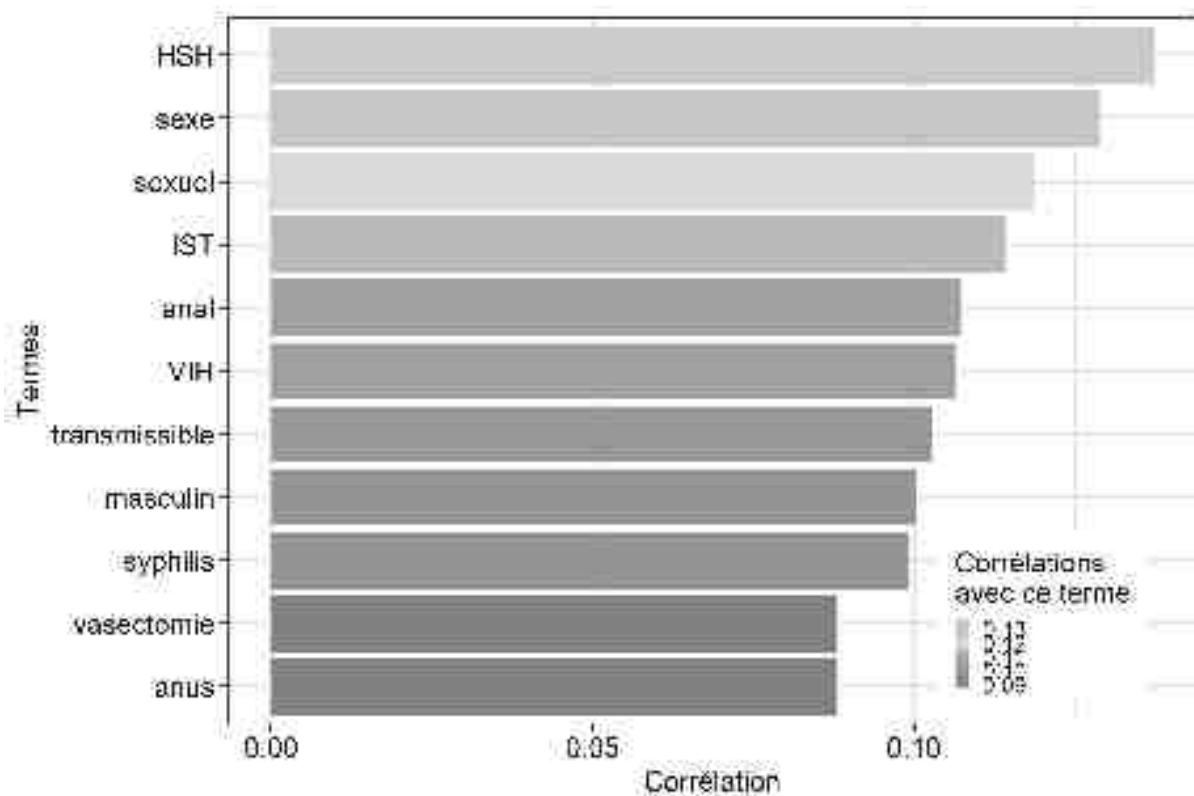
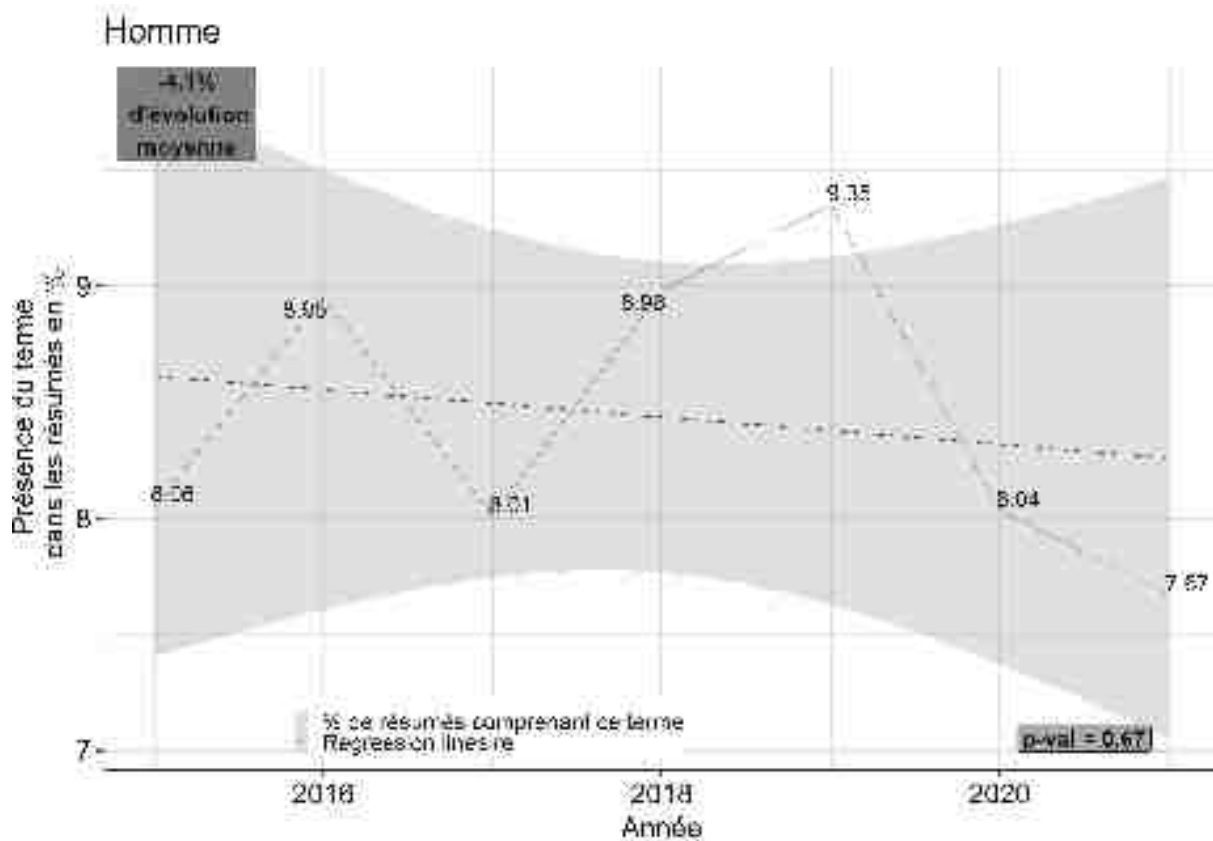
Annexe 24 : Évolution et corrélations du terme prescription sur la période 2015 à 2021



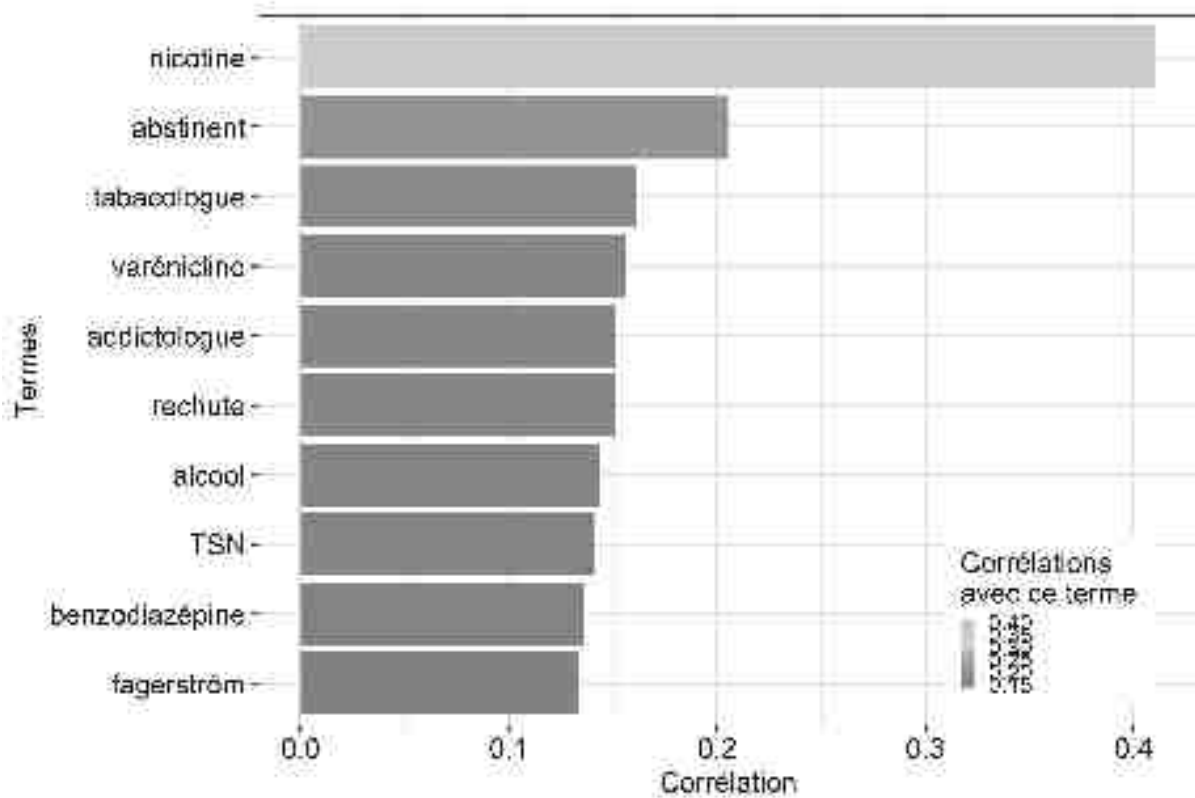
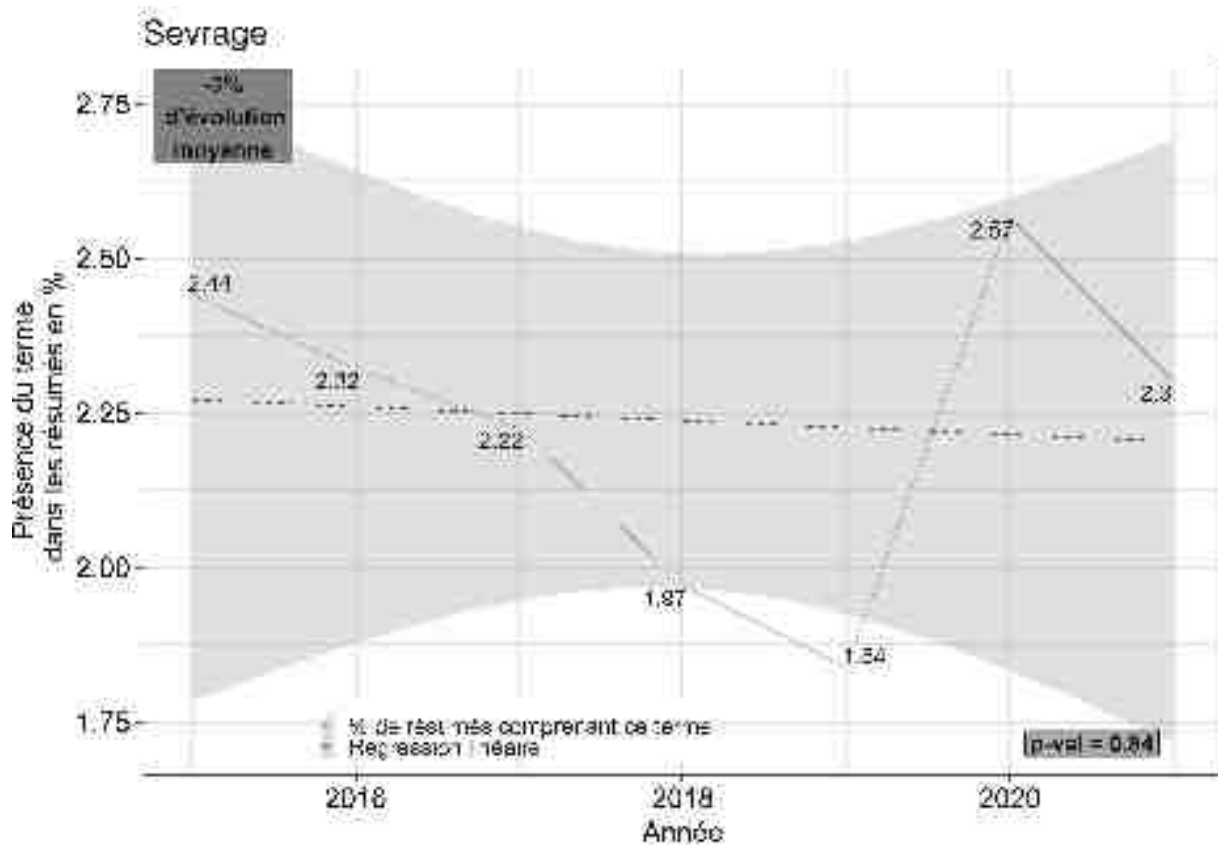
Annexe 25 : Évolution et corrélations du terme palliatif sur la période 2015 à 2021



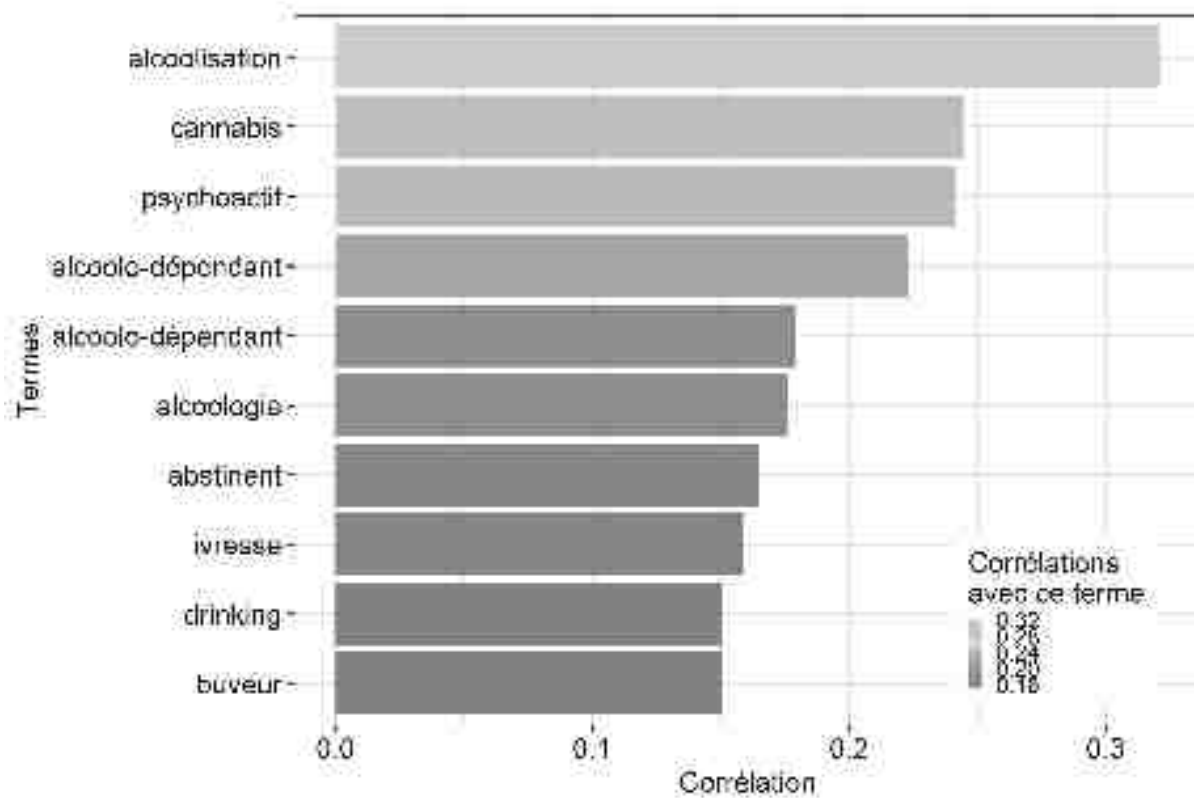
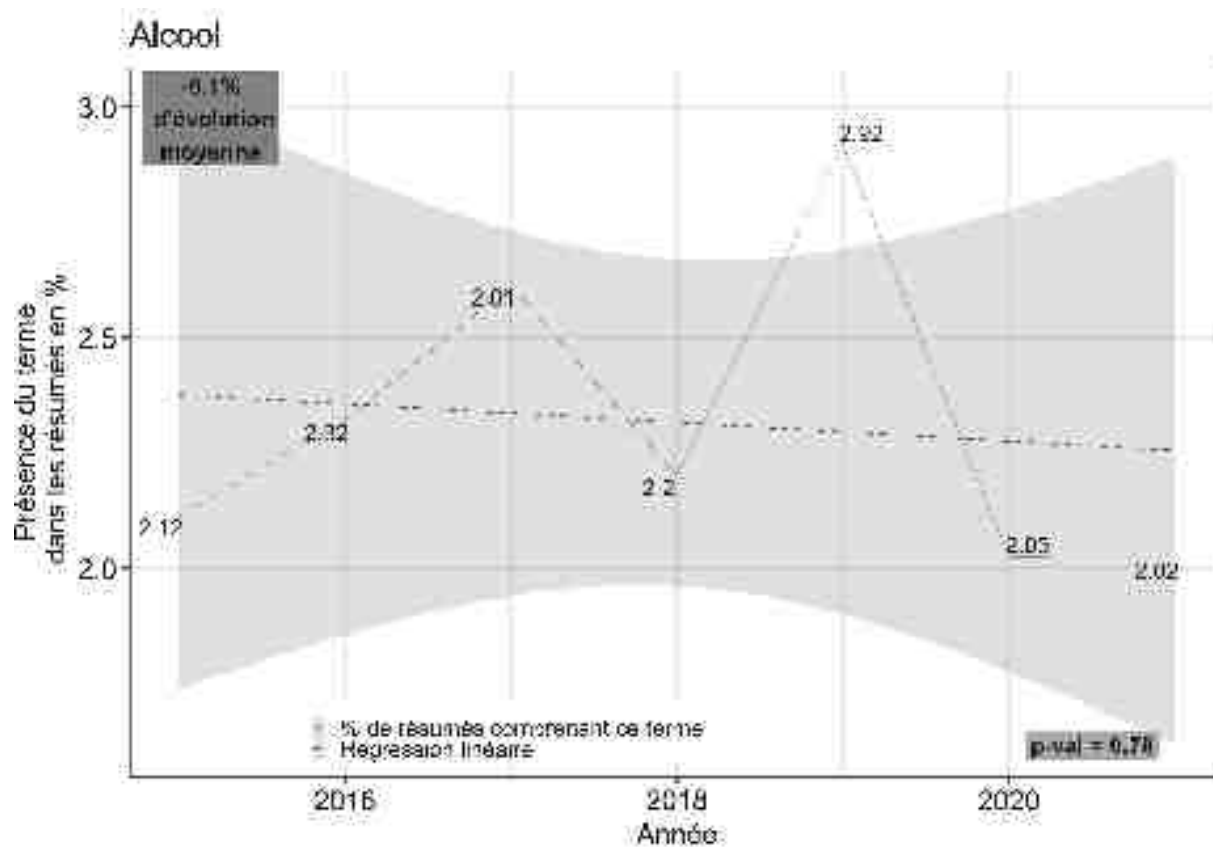
Annexe 26 : Évolution et corrélations du terme homme sur la période 2015 à 2021



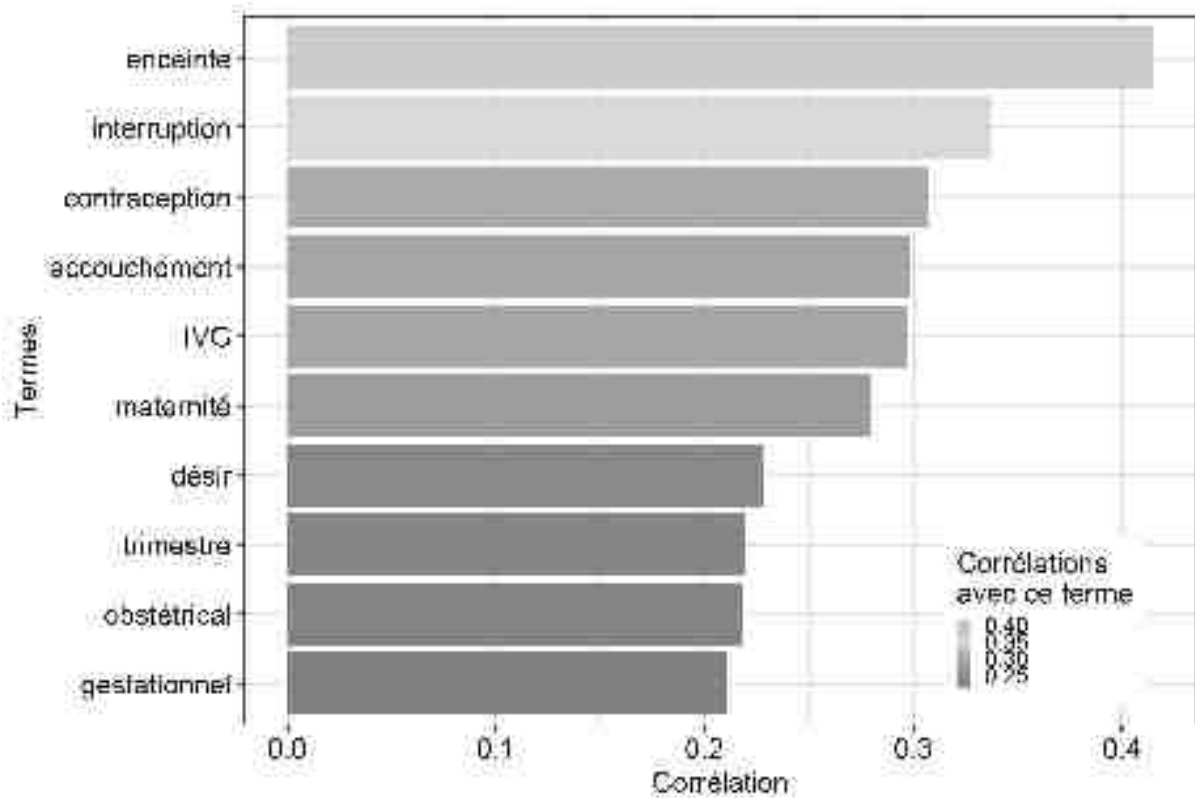
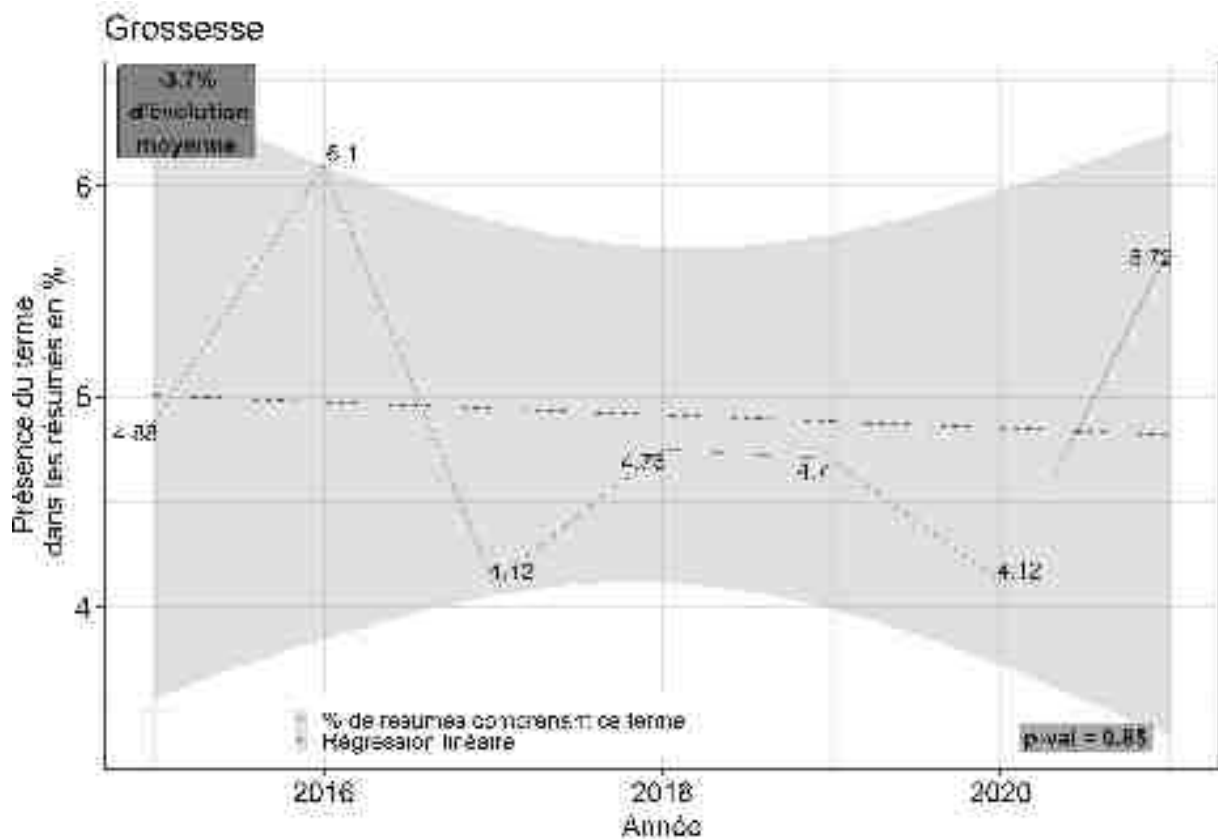
Annexe 27 : Évolution et corrélations du terme sevrage sur la période 2015 à 2021



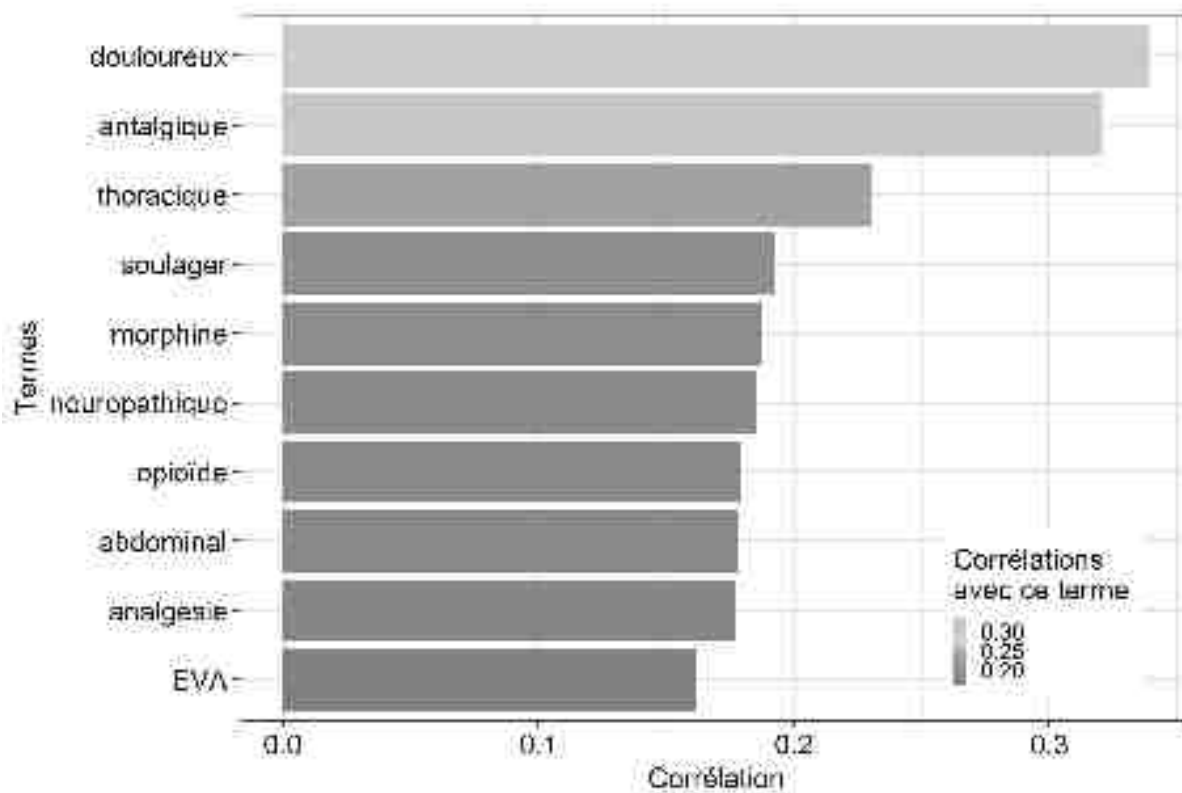
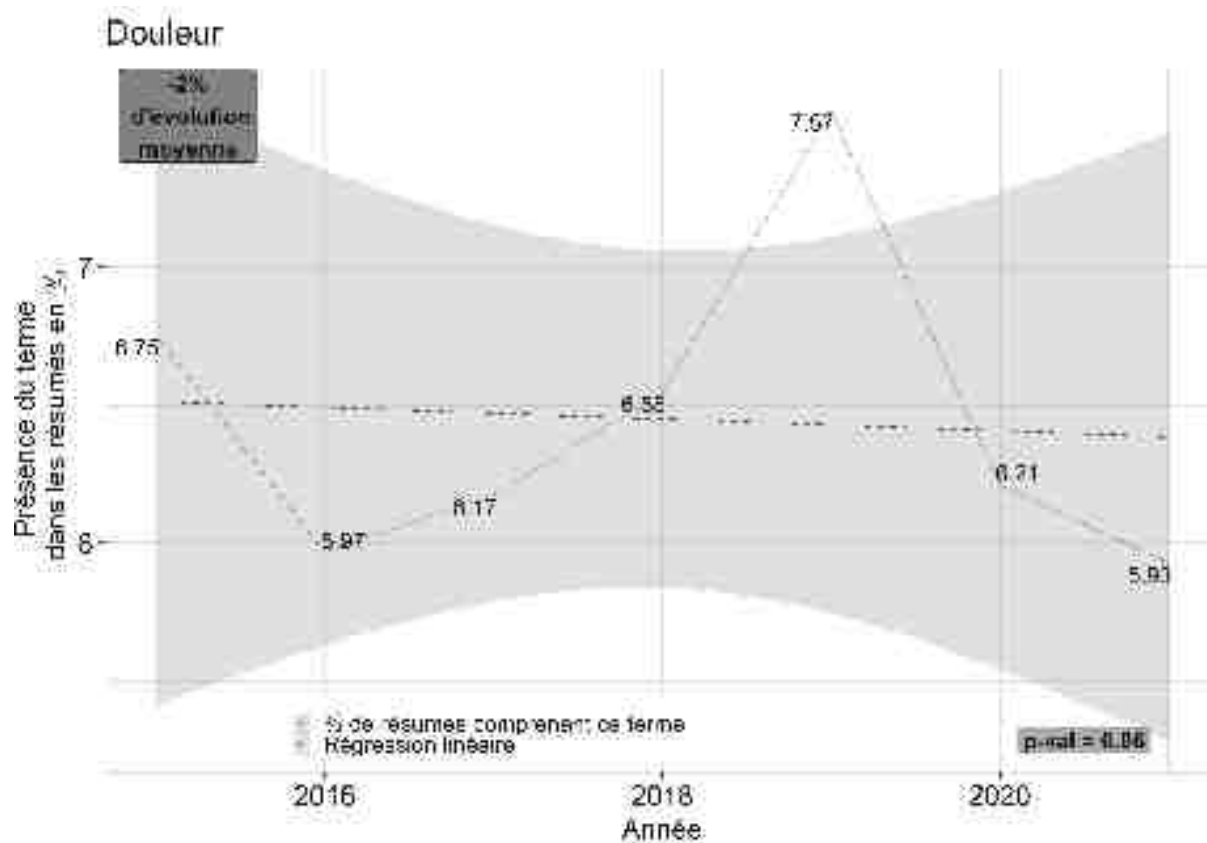
Annexe 28 : Évolution et corrélations du terme alcool sur la période 2015 à 2021



Annexe 29 : Évolution et corrélations du terme grossesse sur la période 2015 à 2021



Annexe 30 : Évolution et corrélations du terme douleur sur la période 2015 à 2021



7 Bibliographie

1. Centre de Ressources Documentaires sur l'Allaitement Maternel [Internet]. Information Pour l'Allaitement. 2021 [cité 11 oct 2022]. Disponible sur: <https://info-allaitement.org/sinformer/service-documentation/>
2. Hauguel, Marie-Catherine (1989-....). Etude EXDEMPAGE : enquête sur la santé et les habitudes de vie de collégiens de la Métropole Européenne de Lille avant une intervention ciblée : aspects méthodologiques et rédaction du protocole [Internet] [Thèse d'exercice]. [Lille]: Médecine : Université Lille 2; 2016 [cité 11 oct 2022]. Disponible sur: https://pepите-depot.univ-lille.fr/LIBRE/Th_Medecine/2016/2016LIL2M241.pdf
3. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ. 6 mai 2000;320(7244):1240.
4. Leycuras L. Recherche allaitement - Lactaclic [Internet]. Lactaclic. 2017 [cité 11 oct 2022]. Disponible sur: <https://lactaclic.fr>
5. Maslach C, Jackson SE, Leiter MP. Maslach Burnout Inventory: Third edition. In: Evaluating stress: A book of resources. Lanham, MD, US: Scarecrow Education; 1997. p. 191-218.
6. Maisonneuve H. Guide pratique du thésard [Internet]. 7ème édition. Paris: Éd. scientifiques L & C [pour] Sanofi Aventis; 2012 [cité 12 août 2021]. 84 p. Disponible sur: <https://docplayer.fr/1473644-Guide-pratique-du-thesard-7-e-edition-pr-herve-maisonneuve.html>
7. Catalogue SUDOC [Internet]. SUDOC. 2022 [cité 13 sept 2022]. Disponible sur: <http://www.sudoc.abes.fr/cbs/xslt/>
8. Chiaro I. Étude STAN-MG : classification des Thèses soutenues À Nancy en Médecine Générale entre 2004 et 2014 [Internet] [Thèse d'exercice]. [NANCY]: FACULTE DE MEDECINE DE NANCY; 2016 [cité 12 sept 2022]. Disponible sur: <https://hal.univ-lorraine.fr/hal-01932322/document>

9. Balva H, Tanguy M, Fanello S, Garnier F. Étude comparative des thèses de médecine générale soutenues à la faculté de médecine d'Angers de 1997 à 1999 et de 2007 à 2009. *Pédagogie Médicale*. août 2012;13(3):159-69.
10. CHARLAS A. Analyse des thèses d'exercice traitant de médecine générale soutenues par les résidents et internes de médecine générale à Saint Etienne de 2003 à 2012 (direction, méthodes, thématiques et valorisation) [Internet] [Thèse d'exercice]. [Saint-Etienne]: faculté de médecine; 2013 [cité 12 août 2021]. Disponible sur: <http://www.sudoc.abes.fr/cbs/xslt/DB=2.1//SRCH?IKT=12&TRM=172556309>
11. Rosenberg C. Analyse des thèses d'exercices soutenues par les internes de médecine générale de Clermont-Ferrand des promotions 2004 à 2016 [Internet] [Thèse d'exercice]. [Clermont-Ferrand]: Clermont Auvergne; 2020 [cité 25 août 2021]. Disponible sur: <https://dumas.ccsd.cnrs.fr/dumas-03248094>
12. Thompson P, Batista-Navarro RT, Kontonatsios G, Carter J, Toon E, McNaught J, et al. Text Mining the History of Medicine. *PLOS ONE*. 6 janv 2016;11(1):33.
13. Rajman M, Besançon R. Text mining: natural language techniques and text mining applications. In: *Data mining and reverse engineering*. Springer; 1998. p. 50-64.
14. KANNAN KG. TOPCAT: DATA MINING FOR TOPIC IDENTIFICATION IN A TEXT CORPUS. 1956;
15. Moore GE. Cramming more components onto integrated circuits. 1965;
16. Moulinier L. Un aspect particulier de la médecine des religieux après le XIIe siècle: l'attrait pour l'astrologie médicale. In *Publications de l'École française de Rome*; 2010. p. 59-86.
17. De Rood M. Histoire de l'anesthésie en Belgique. *Revue médicale de Bruxelles*. 2012;33(3):179-87.

18. Décret n°97-495 du 16 mai 1997 relatif au stage pratique des résidents auprès des praticiens généralistes agréés. 97-495 mai 16, 1997.
19. Décret n° 83-691 du 26 juillet 1983 fixant le rôle, la composition et les modalités de fonctionnement des commissions prévues aux articles 57 et 60 de la loi n° 68-978 du 12 novembre 1968 modifiée d'orientation de l'enseignement supérieur. - Légifrance [Internet]. [cité 11 janv 2023]. Disponible sur: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000686331>
20. Historique du CNGE [Internet]. 2012 [cité 11 janv 2023]. Disponible sur: https://www.cnge.fr/le_cnge/historique_du_cnge/
21. Arrêté du 22 septembre 2004 fixant la liste et la réglementation des diplômes d'études spécialisées de médecine.
22. Décret n° 2008-744 du 28 juillet 2008 portant dispositions relatives aux personnels enseignants des universités, titulaires et non titulaires de médecine générale. 2008-744 juill 28, 2008.
23. Europe W. The European definition of general practice/family medicine. Barcelona: WONCA Europe. 2002;
24. Vernaudo D. Analyse thématique des thèses d'exercice soutenues par les internes de médecine générale de Strasbourg en 2016 et 2017 [Internet] [Thèse d'exercice]. [2009-...., France]: Université de Strasbourg; 2019 [cité 3 nov 2022]. Disponible sur: https://publication-theses.unistra.fr/public/theses_exercice/MED/2019/2019_VERNAUDON_Damien.pdf
25. White KL, Williams TF, Greenberg BG. The ecology of medical care. 1961. Bulletin of the New York Academy of Medicine. 1996;73(1):187.
26. Green LA, Fryer Jr GE, Yawn BP, Lanier D, Dovey SM. The ecology of medical care revisited. New England Journal of Medicine. 2001;344(26):2021-5.

27. Coulhon T, Sachwald F. La position scientifique de la France dans le monde et en Europe. L'ENA hors les murs. 2021;509(8):18-22.
28. Lionis C, Stoffers H, Hummers-Pradier E, Griffiths F, Rotar-Pavlič D, Rethans J. Setting priorities and identifying barriers for general practice research in Europe. Results from an EGPRW* meeting. Family Practice. 2004;21(5):587-93.
29. Qu'est que le DMG de l'Université de Paris | Département Médecine Générale - Université Paris Cité [Internet]. [cité 11 janv 2023]. Disponible sur: <https://dmg-u-paris.fr/c/le-departement>
30. Circuit de labellisation CNGE des projets de recherche [Internet]. [cité 11 janv 2023]. Disponible sur: https://www.cnge.fr/conseil_scientifique/circuit_de_labellisation_cnge_des_projets_de_reche/
31. Société Française de Médecine Générale : Présentation [Internet]. [cité 11 janv 2023]. Disponible sur: <https://www.sfm.org/presentation/>
32. SFDRMG - Société Française de Documentation et Recherche [Internet]. SFDRMG. [cité 11 janv 2023]. Disponible sur: <https://www.sfdrmg.org/societe-francaise-de-documentation-et-recherche-en-medecine-generale/>
33. Recherche [Internet]. [cité 11 janv 2023]. Disponible sur: <https://www.sftg.eu/recherche/>
34. Résultats des appels à projet - ANSM : Agence nationale de sécurité du médicament et des produits de santé [Internet]. [cité 11 janv 2023]. Disponible sur: [https://archiveansm.integra.fr/L-ANSM/Appels-a-projets-Associations/Resultats-des-appels-a-projet/\(offset\)/0](https://archiveansm.integra.fr/L-ANSM/Appels-a-projets-Associations/Resultats-des-appels-a-projet/(offset)/0)
35. appelsprojetsrecherche.fr - Bienvenue [Internet]. [cité 11 janv 2023]. Disponible sur: <https://www.appelsprojetsrecherche.fr/>
36. Décret n°2004-67 du 16 janvier 2004 relatif à l'organisation du troisième cycle des études médicales. 2004-67 janv 16, 2004.

37. DGOS_Michel.C, DGOS_Michel.C. La réforme du 3ème cycle des études de médecine [Internet]. Ministère de la Santé et de la Prévention. 2023 [cité 11 janv 2023]. Disponible sur: <https://solidarites-sante.gouv.fr/professionnels/se-former-s-installer-exercer/etudes-de-medecine-reforme-du-3eme-cycle/r3c>
38. Attali C, Bail P, Magnier AM, Beis JN, Ghasarossian C, Gomes J, et al. Compétences pour le DES de médecine générale.
39. Arrêté du 12 avril 2017 portant organisation du troisième cycle des études de médecine.
40. dit RM. Validation du 3ème cycle [Internet]. ISNI. [cité 11 janv 2023]. Disponible sur: <https://isni.fr/validation-du-3eme-cycle/>
41. LOI n° 2022-1616 du 23 décembre 2022 de financement de la sécurité sociale pour 2023 (1). 2022-1616 déc 23, 2022.
42. Levasseur G, Schweyer FX. La recherche en médecine générale, à travers les thèses de médecine. *Sante Publique*. 2003;Vol. 15(2):203-12.
43. Balva H, Tanguy M, Fanello S, Garnier F. Étude comparative des thèses de médecine générale soutenues à la faculté de médecine d'Angers de 1997 à 1999 et de 2007 à 2009. *Pédagogie Médicale*. 1 août 2012;13(3):159-69.
44. Mathieu L. Analyse des thèses d'exercice soutenues par les internes en médecine générale de Lyon de Janvier 2011 à Décembre 2013 [Thèse d'exercice]. [Lyon]: Université Claude Bernard Lyon 1; 2015.
45. Briquet Y. Infection urinaire de l'adulte: prise en charge par les médecins généralistes en Guyane Française [Internet] [Thèse d'exercice]. [Amiens]: Université de Picardie Jules Verne; 2016 [cité 3 nov 2022]. Disponible sur: <https://dumas.ccsd.cnrs.fr/dumas-01380849/document>

46. Qaiser S, Ali R. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *IJCA*. 16 juill 2018;181(1):5.
47. Arun R, Suresh V, Veni Madhavan CE, Narasimha Murthy MN. On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In: Zaki MJ, Yu JX, Ravindran B, Pudi V, éditeurs. *Advances in Knowledge Discovery and Data Mining*. Berlin, Heidelberg: Springer; 2010. p. 391-402. (Lecture Notes in Computer Science; vol. 6118).
48. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. *Neurocomputing*. 1 mars 2009;72(7):1775-81.
49. Steyvers M, Smyth P, Rosen-Zvi M, Griffiths T. Probabilistic author-topic models for information discovery. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining [Internet]*. New York, NY, USA: Association for Computing Machinery; 2004 [cité 11 oct 2022]. p. 306-15. (KDD '04). Disponible sur: <https://doi.org/10.1145/1014052.1014087>
50. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*. 2014;17(1):61-84.
51. Blei DM. Probabilistic topic models. *Commun ACM*. 1 avr 2012;55(4):77-84.
52. Arrêté du 12 avril 2017 portant organisation du troisième cycle des études de médecine [Internet]. MENS1708241A, 0089 avr 14, 2017. Disponible sur: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000034419758/>
53. Statistiques ECN : Médecine générale [Internet]. [cité 11 oct 2022]. Disponible sur: <https://www.medshake.net/medecine/ECN/statistiques/medecine-generale/>
54. Reynaud T, Geffroy S. D'une PASS de ville aux soins primaires : expériences des patients pris en charge par le dispositif « PASS de Ville » de Médecins du Monde [Internet] [médecine générale].

[marseille]: faculté de médecine; 2021 [cité 24 sept 2022]. Disponible sur: <https://dumas.ccsd.cnrs.fr/dumas-03591100>

55. Pennacchi C. État des lieux de la gestion à domicile par les médecins généralistes de la Somme des symptômes qui caractérisent les évolutions des pathologies qui affectent les patients palliatifs [Thèse d'exercice]. [France]: Université de Picardie Jules Verne; 2021.

56. Helfer A. Effets et conséquences du confinement lié à la pandémie de COVID-19 sur l'activité physique chez des patients atteints de pathologies neurologiques chroniques (Sclérose en plaques, Maladie de Parkinson et Fibromyalgie): Étude PhysiCOVID [Internet] [Thèse d'exercice]. [2009-...., France]: Université de Strasbourg; 2020 [cité 18 nov 2022]. Disponible sur: https://publication-theses.unistra.fr/restreint/theses_exercice/MED/2020/2020_HELFER_Amandine.pdf

57. Loi n° 2004-810 du 13 août 2004 relative à l'assurance maladie [Internet]. 420 juill 30, 2004. Disponible sur: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000625158/>

58. Qu'est-ce-que le parcours de soins ? - Ministère de la Santé et de la Prévention [Internet]. [cité 15 oct 2022]. Disponible sur: <https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/parcours-de-sante-vos-droits/liberte-de-choix-et-acces-aux-soins/article/qu-est-ce-que-le-parcours-de-soins>

59. Sivignon F, Simonnot N, Bret C, Fanget D, Reboul C. L'accueil sanitaire des migrants en France et en Europe, reflet d'une crise de solidarité. Bulletin de l'Académie Nationale de Médecine. 2019;203(1):33-9.

60. Ramazzini B. Essai sur les maladies des artisans. Chez Moutard, Imprimeur-Libraire de la Reine, de Madame, & de Madame la Comtesse d'Artois, rue du Hurepoix, près le Pont Saint-Michel; 1777. 670 p.

61. Gelly M, Pitti L. Une médecine de classe ? Inégalités sociales, système de santé et pratiques de soins. Agone. 2016;58(1):7-18.

62. Raude J. L'hésitation vaccinale : une perspective psychosociologique. Bulletin de l'Académie Nationale de Médecine. 1 févr 2016;200(2):199-209.
63. Marre-Fournier F. Le programme national nutrition santé 4, une (r)évolution dans nos assiettes. Actualités Pharmaceutiques. 1 janv 2021;60(602):18-22.
64. Bloy G, Moussard Philippon L, Rigal L. Les médecins généralistes et le conseil en activité physique : des évidences aux contingences de la consultation. Santé Publique. 2016;S1(HS):153-61.
65. Chantry AA, Monier I, Marcellin L. Allaitement maternel (partie 1) : fréquence, bénéfices et inconvénients, durée optimale et facteurs influençant son initiation et sa prolongation. Recommandations pour la pratique clinique. Journal de Gynécologie Obstétrique et Biologie de la Reproduction. 1 déc 2015;44(10):1071-9.
66. Vigoureux S, Le Guen M. Contexte de la contraception en France. RPC Contraception CNGOF. Gynécologie Obstétrique Fertilité & Sénologie. 1 déc 2018;46(12):777-85.
67. Jedat V, Desnouhes A, Andrieux M, Besnier M, Archambault P. État des lieux des actions favorisant l'installation des médecins généralistes en France métropolitaine. Santé Publique. 2022;34(2):231-41.
68. Les motifs de recours hospitalier en court séjour en 2018 et leurs évolutions récentes [Internet]. [cité 16 oct 2022]. Disponible sur: <https://drees.solidarites-sante.gouv.fr/sites/default/files/2020-12/er1168.pdf>
69. Bagein G, Costemalle V, Deroyon T, Hazo JB, Naouri D, Pesonel E, et al. L'état de santé de la population en France. Les dossiers de la DREES. 2022;(n° 102):70.
70. Eco-Santé France, Régions & Départements [Internet]. [cité 17 oct 2022]. Disponible sur: <http://www.ecosante.fr/index2.php?base=DEPA&langs=FRA&langh=FRA&source=357>

71. Colin A. État des lieux de la visite à domicile par les médecins généralistes, en France, en 2016 [Internet] [Thèse d'exercice]. [PARIS]: Faculté de Médecine PARIS DESCARTES; 2019. Disponible sur: <https://dumas.ccsd.cnrs.fr/dumas-02505145>
72. Diplôme de formation approfondie en sciences médicales (DFASM) | SMPM [Internet]. [cité 11 janv 2023]. Disponible sur: <https://smpm.univ-amu.fr/fr/formations/medecine/diplome-formation-approfondie-sciences-medicales-dfasm>
73. 2021-Programme-DES-MG.pdf [Internet]. [cité 11 janv 2023]. Disponible sur: <https://medphar.univ-poitiers.fr/wp-content/uploads/sites/169/2021/09/2021-Programme-DES-MG.pdf>
74. Ségur M. La 5G, à la croisée des chemins. *Futuribles*. 2021;443(4):86-90.
75. 94 % des 15-29 ans ont un smartphone en 2021 - Insee Focus - 259 [Internet]. [cité 14 janv 2023]. Disponible sur: <https://www.insee.fr/fr/statistiques/6036909>
76. Téléconsultation [Internet]. [cité 14 janv 2023]. Disponible sur: <https://www.ameli.fr/exercice-coordonne/exercice-professionnel/telemedecine/teleconsultation>
77. Monziols M, Chaput H, Verger P, Scronias D, Ventelou B, Barlet M. Trois médecins généralistes sur quatre ont mis en place la téléconsultation depuis le début de l'épidémie de Covid-19. *DREES Études et résultats*. 2020;1162:4.
78. Tristan M. Enquête sur les délais de rendez-vous pour soins non programmés en médecine générale libérale en Sarthe. 2018;
79. Karsenty J; VP Isabelle. Utilisation des écrans avant 6 ans expériences des parents à travers une étude qualitative perspectives d'accompagnement en médecine générale [Internet]. 2020. Disponible sur: <https://dumas.ccsd.cnrs.fr/dumas-03121608>

80. Société canadienne de pédiatrie groupe de travail sur la santé numérique Ottawa (Ontario). Le temps d'écran et les jeunes enfants : promouvoir la santé et le développement dans un monde numérique. *Paediatrics & Child Health*. 27 nov 2017;22(8):469-77.
81. Brun B; MB David ; Meurice, Jean Claude ; Parthenay, Pascal ; Garcia, Magali. Etude des modifications de pratiques des médecins généralistes charentais face à la COVID 19 [Internet]. 2020. Disponible sur: <http://nuxeo.edel.univ-poitiers.fr/nuxeo/site/esupversions/9ffa9337-4281-419f-8e6b-e4118d30ec33>
82. LOI n° 2019-1480 du 28 décembre 2019 visant à agir contre les violences au sein de la famille (1). 2019-1480 déc 28, 2019.
83. LOI n° 2022-52 du 24 janvier 2022 relative à la responsabilité pénale et à la sécurité intérieure (1). 2022-52 janv 24, 2022.
84. LOI n° 2022-52 du 24 janvier 2022 relative à la responsabilité pénale et à la sécurité intérieure (1). 2022-52 janv 24, 2022.
85. Sánchez OR, Vale DB, Rodrigues L, Surita FG. Violence against women during the COVID-19 pandemic: An integrative review. *International Journal of Gynecology & Obstetrics*. 2020;151(2):180-7.
86. Vallot M; A Caroline. Difficulté d'accès aux soins pour les patients dépressifs évaluation de la prise en charge par le médecin généraliste [Internet]. 2020. Disponible sur: <https://dumas.ccsd.cnrs.fr/dumas-02962377>
87. Donnet M. 14. La démographie médicale et les déserts médicaux. In: *Doctor's blues ou le burnout des médecins*. EDP Sciences; 2021. p. 51-60.
88. Gakunzi S Kiyago, Johanna ; Dardel, Nicole. Le suivi de grossesse en situation de précarité vécu et pratiques de professionnels de soins primaires. 2020.
89. Plotkin SL, Plotkin SA. A short history of vaccination. *Vaccines*. 2004;5:1-16.

90. Tableau des vaccins existants en France | Vaccination Info Service [Internet]. [cité 4 janv 2023]. Disponible sur: https://vaccination-info-service.fr/Les-vaccins-existants-en-France/Tableau-des-vaccins-existants-en-France?gclid=EAlaIqObChMIq5-6nau_AIVBY9oCR3roQ6GEAMYASAAEgLPB_D_BwE&gclsrc=aw.ds
91. Plotkin SA, Orenstein WA. Vaccines. Elsevier Health Sciences; 2008. 1748 p.
92. Vaccins et vaccination [Internet]. [cité 17 janv 2023]. Disponible sur: <https://www.who.int/fr/health-topics/vaccines-and-immunization>
93. Fonteneau L, Barret AS, Lévy-Bruh D. Évolution de la couverture vaccinale du vaccin contre le papillomavirus en France - 2008-2018 (*). 2020;(354).
94. Wyplosz B, Fernandes J, Raguideau F, Leboucher C, Goussiaume G, Moïsi J. COVARISQ (estimation de la COuverture VAccinale des adultes à RISQues): taux de vaccination des malades atteints de comorbidités en France en 2017. Médecine et Maladies Infectieuses. 2020;50(6):S179-80.
95. Martinez Caballero M de los A. Evaluation de l'utilisation de la méthode cocooning pour la prévention de la coqueluche chez le nourrisson en médecine générale. 2020.
96. DeStefano F, Weintraub ES, Chen RT. Recombinant hepatitis B vaccine and the risk of multiple sclerosis: a prospective study. 2022;
97. Maisonneuve H, Floret D. Affaire Wakefield: 12 ans d'errance car aucun lien entre autisme et vaccination ROR n'a été montré. La presse médicale. 2012;41(9):827-34.
98. Décret n° 2016-1990 du 30 décembre 2016 relatif aux conditions de dispensation de l'activité physique adaptée prescrite par le médecin traitant à des patients atteints d'une affection de longue durée [Internet]. AFSP1637993D déc 31, 2016. Disponible sur: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000033748987/2023-01-17/>

99. Wagner S, Kersuzan C, Gojard S, Tichit C, Nicklaus S, Geay B, et al. Durée de l'allaitement en France selon les caractéristiques des parents et de la naissance. Résultats de l'étude longitudinale française Elfe, 2011. Bulletin Epidémiologique Hebdomadaire-BEH. 2015;(27):522-32.
100. Salanave B, De Launay C, Guerrisi C, Castetbon K. Taux d'allaitement maternel à la maternité et au premier mois de l'enfant. Résultats de l'étude Épipane, France, 2012. J Pédiatrie Puériculture. 2012;25(6):364-72.
101. Arrêté du 27 novembre 2017 modifiant l'arrêté du 12 avril 2017 relatif à l'organisation du troisième cycle des études de médecine et l'arrêté du 21 avril 2017 relatif aux connaissances, aux compétences et aux maquettes de formation des diplômes d'études spécialisées et fixant la liste de ces diplômes et des options et formations spécialisées transversales du troisième cycle des études de médecine [Internet]. nov 27, 2017. Disponible sur: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000036237037>
102. Certains préservatifs sont gratuits en pharmacie pour les moins de 26 ans dès janvier 2023 [Internet]. [cité 17 janv 2023]. Disponible sur: <https://www.service-public.fr/particuliers/actualites/A16208>
103. Contraception gratuite pour les femmes de moins de 26 ans depuis le 1er janvier 2022 [Internet]. [cité 17 janv 2023]. Disponible sur: <https://www.service-public.fr/particuliers/actualites/A15158>
104. Domerc A; C Barbara. Intérêt d'une brochure d'information sur le dépistage du cancer du col de l'utérus dans la connaissance du dépistage par frottis cervico-utérin évaluation à 2 mois. 2020.
105. Santé publique France - YouTube [Internet]. [cité 17 janv 2023]. Disponible sur: <https://www.youtube.com/@SantepubliqueFrance>
106. Ohayon S; B Mahrez. Les freins à la vaccination antipneumococcique étude du patient âgé à risque en cabinet de médecine générale. 2020.

107. Ramirez A, Bascou A, Vergnault M, Telmon N, Savall F. Caractéristiques des violences répétées observées sur 10 années consécutives d'une consultation de médecine légale du vivant. *La Revue de Médecine Légale*. 2022;13(1):17-22.
108. Peyron PA, David M. [Suicide risk assessment tools for adults in general medical practice]. *Presse Medicale* (Paris, France: 1983). juin 2015;44(6 Pt 1):590-600.
109. Colquitt JL, Loveman E, O'Malley C, Azevedo LB, Mead E, Al-Khudairy L, et al. Diet, physical activity, and behavioural interventions for the treatment of overweight or obesity in preschool children up to the age of 6 years. *Cochrane Database of Systematic Reviews* [Internet]. 2016 [cité 16 mai 2019];(3). Disponible sur: <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD012105/full/fr?highlightAbstract=withdrawn%7Cobes%7Cpsychology%7Cobesity%7Cpsycholog>
110. DGOS_Michel.C, DGOS_Michel.C. Le principe du CESP [Internet]. Ministère de la Santé et de la Prévention. 2023 [cité 17 janv 2023]. Disponible sur: <https://solidarites-sante.gouv.fr/professionnels/se-former-s-installer-exercer/le-contrat-d-engagement-de-service-public-cesp/article/le-principe-du-cesp>
111. LOI n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé. 2016-41 janv 26, 2016.
112. Aide à l'installation ou à la pratique du médecin en zone sous-dotée [Internet]. [cité 17 janv 2023]. Disponible sur: <https://www.ameli.fr/medecin/exercice-liberal/vie-cabinet/aides-financieres/pratique-zones-sous-dotees>
113. CAB_Solidarites, CAB_Solidarites. Feuille de route - Santé mentale et psychiatrie - Jeudi 28 juin 2018 [Internet]. Ministère de la Santé et de la Prévention. 2023 [cité 17 janv 2023]. Disponible sur: <https://solidarites-sante.gouv.fr/archives/archives-presse/archives-dossiers-de-presse/article/feuille-de-route-sante-mentale-et-psychiatrie-jeudi-28-juin-2018>

114. Hôpitaux et cliniques : les chiffres clés de l'année 2019 [Internet]. vie-publique.fr. [cité 17 janv 2023]. Disponible sur: <https://www.vie-publique.fr/en-bref/281507-hopitaux-et-cliniques-les-chiffres-cles-de-lannee-2019>
115. Charmarkeh H. Les personnes âgées et la fracture numérique de «second degré»: l'apport de la perspective critique en communication. *Revue française des sciences de l'information et de la communication*. 2015;(6).
116. Garczynski G. Fracture numérique, fracture sociale. *Revue Projet*. 2019;371(4):33-6.
117. Michaud M, Michaud Peyrot C. Réglementation de la recherche médicale en France. *La Revue de Médecine Interne*. 1 févr 2020;41(2):98-105.
118. Série 000442588 Prix moyens mensuels de vente au détail en métropole - Gazole (1 litre) | Insee [Internet]. [cité 17 janv 2023]. Disponible sur: <https://www.insee.fr/fr/statistiques/serie/000442588#Graphique>
119. Le diabète en France : les chiffres 2020 [Internet]. [cité 17 janv 2023]. Disponible sur: </les-actualites/2021/le-diabete-en-france-les-chiffres-2020>
120. PRADAXA 75mg, gélule - PRADAXA 110mg, gélule - PRADAXA 150mg, gélule - ANSM : Agence nationale de sécurité du médicament et des produits de santé [Internet]. [cité 17 janv 2023]. Disponible sur: <https://archiveansm.integra.fr/Activites/Surveillance-des-medicaments/Medicaments-faisant-l-objet-d-un-plan-de-gestion-des-risques/Medicaments-faisant-l-objet-d-un-Plan-de-Gestion-des-Risques-PGR2/PRADAXA-75mg-gelule-PRADAXA-110mg-gelule-PRADAXA-150mg-gelule>
121. XARELTO 10 mg comprimé pelliculé - XARELTO 15 mg comprimé pelliculé - XARELTO 20 mg comprimé pelliculé - ANSM : Agence nationale de sécurité du médicament et des produits de santé [Internet]. [cité 17 janv 2023]. Disponible sur: <https://archiveansm.integra.fr/Activites/Surveillance-des-medicaments/Medicaments-faisant-l-objet-d-un-plan-de-gestion-des-risques/Medicaments->

faisant-l-objet-d-un-Plan-de-Gestion-des-Risques-PGR2/XARELTO-10-mg-comprime-pellicule-
XARELTO-15-mg-comprime-pellicule-XARELTO-20-mg-comprime-pellicule

122. HAS. COMMISSION DE LA TRANSPARENCE Avis eliquis [Internet]. 2018 janv p. 5. Disponible sur: https://www.has-sante.fr/upload/docs/application/pdf/2018-02/eliquis_avis_ct1569916066.pdf

123. Conseil national de l'Ordre des médecins. En France, les étudiants en médecine sont formés au sein des Unités de formation et de recherche (UFR) d'universités associées à l'un des 29 CHU. 21 mai 2019 [cité 7 nov 2022]; Disponible sur: <https://www.conseil-national.medecin.fr/etudiants-internes/etudes-medecine/facultes-medecine>

124. Genin G, Agrifoglio G. Les travaux de thèse en médecine générale, une discipline centrée patient, un travail centré médecin. 2017;52.

125. Labonde B, Plotton C, Charles R, Gocko X. Directions, contenus, valorisations des thèses de médecine générale. Médecine. 2019;15(9):427-32.

126. Vernaoudon D. Analyse thématique des thèses d'exercice soutenues par les internes de Médecine Générale de Strasbourg en 2016 et 2017. [Internet] [Thèse d'exercice]. [Strasbourg]: Université de Strasbourg; 2019 [cité 3 nov 2022]. Disponible sur: https://publication-theses.unistra.fr/public/theses_exercice/MED/2019/2019_VERNAUDON_Damien.pdf

127. CHARLAS A. Analyse des thèses d'exercice traitant de médecine générale soutenues par les résidents et internes de médecine générale à Saint Etienne de 2003 à 2012 (direction, méthodes, thématiques et valorisation) [Internet] [Thèse d'exercice]. [Saint-Etienne]: faculté de médecine; 2013 [cité 12 août 2021]. Disponible sur: <http://www.sudoc.abes.fr/cbs/xslt/DB=2.1//SRCH?IKT=12&TRM=172556309>

128. Wonca Europe [Internet]. [cité 21 oct 2022]. Disponible sur: https://www.cnge.fr/l_international/wonca_europe/

129. Mots clé MeSH et CISMef | Département Médecine Générale - Université Paris Cité [Internet]. [cité 21 oct 2022]. Disponible sur: <https://dmg-u-paris.fr/p/mots-cle-mesh-et-cismef>
130. International Classification of Primary Care, 2nd edition (ICPC-2) [Internet]. [cité 21 oct 2022]. Disponible sur: <https://www.who.int/standards/classifications/other-classifications/international-classification-of-primary-care>
131. Connolly JP, McGavock H, Wilson-Davis K. Research methodology: coding perceived morbidity in general practice—an evaluation of the read classification and the international classification of primary care (ICPC). *Pharmacoepidemiology and Drug Safety*. 1997;6(5):325-30.
132. Frese T, Herrmann K, Bungert-Kahl P, Sandholzer H. Inter-rater reliability of the ICPC-2 in a German general practice setting. *Swiss Medical Weekly* [Internet]. 12 août 2012 [cité 21 oct 2022];(33). Disponible sur: <https://smw.ch/article/doi/smw.2012.13621>

DÉCLARATION SUR L'HONNEUR



Faculté de médecine
maieutique et sciences de la santé
Université de Strasbourg

Document avec signature originale devant être joint :

- à votre mémoire de D.E.S.
- à votre dossier de demande de soutenance de thèse

Nom : ZOLFER Prénom : MICHAËL

Ayant été informé(e) qu'en m'appropriant tout ou partie d'une œuvre pour l'intégrer dans mon propre mémoire de spécialité ou dans mon mémoire de thèse de docteur en médecine, je me rendrais coupable d'un délit de contrefaçon au sens de l'article L335-1 et suivants du code de la propriété intellectuelle et que ce délit était constitutif d'une fraude pouvant donner lieu à des poursuites pénales conformément à la loi du 23 décembre 1901 dite de répression des fraudes dans les examens et concours publics,

Ayant été avisé(e) que le président de l'université sera informé de cette tentative de fraude ou de plagiat, afin qu'il saisisse la juridiction disciplinaire compétente,

Ayant été informé(e) qu'en cas de plagiat, la soutenance du mémoire de spécialité et/ou de la thèse de médecine sera alors automatiquement annulée, dans l'attente de la décision que prendra la juridiction disciplinaire de l'université.

J'atteste sur l'honneur

Ne pas avoir reproduit dans mes documents tout ou partie d'œuvre(s) déjà existante(s), à l'exception de quelques brèves citations dans le texte, mises entre guillemets et référencées dans la bibliographie de mon mémoire.

A écrire à la main : « J'atteste sur l'honneur avoir connaissance des suites disciplinaires ou pénales que j'encours en cas de déclaration erronée ou incomplète ».

J'atteste sur l'honneur avoir connaissance des suites disciplinaires ou pénales que j'encours en cas de déclaration erronée ou incomplète

Signature originale :

A WALSCHETD, le 27/10/2022

Photocopie de cette déclaration devant être annexée en dernière page de votre mémoire de D.E.S. ou de Thèse.

RÉSUMÉ :

Introduction : En médecine générale, la thèse concerne plus de 3000 doctorants annuellement. Chaque thèse est propre à chaque interne, mais y a-t-il dans tous ces travaux des éléments comparables pouvant mettre en exergue des besoins spécifiques dans la population générale ou en matière de formation initiale ?

Méthode : Cette étude porte sur les résumés de thèses de médecine générale de France de 2015 à 2021 disponibles sur la plateforme Sudoc. Après la recherche, l'extraction, le tri et l'uniformisation des données, les résultats ont été obtenus via des algorithmes de fouille de texte ou text mining.

Résultats : Au total, 11 026 fiches de thèses ont été analysées. La période d'étude s'étend de 2015 à 2021. Les principales villes représentées sont Paris à 14,22 %, Lille à 12,84%, Bordeaux à 12,11% et Marseille à 8,74%. Les thèmes qui sont les plus fréquents sont la formation des médecins généralistes (3,7%), les pratiques en médecine générale (3,4%), et la relation médecin-patient (3,2%). Les thèses mentionnant la téléconsultation, des écrans, du COVID et des violences ainsi que les thèmes sur l'outil informatique, la vaccination, et la santé mentale sont en forte augmentation. Les thèses mentionnant des hospitalisations, des personnes âgées et des soins à domicile, ainsi que les thèmes sur les urgences les hospitalisations, la personne âgée fragile et les anticoagulants sont en forte diminution.

Discussion : Cette analyse peut être réactualisée annuellement. Elle pourrait être la base d'un outil permettant de guider les futurs travaux de thèse. En supposant que les internes choisissent leur sujet de thèse en fonction des potentiels manques de connaissances qu'ils ont, cette analyse permet également de mettre en évidence des thématiques importantes à approfondir pendant le cursus médical.

Rubrique de classement : médecine générale

Mots-clés :

- Dissertation universitaire
 - Médecine générale
 - Fouille de données
 - Dissertations universitaires comme sujet
 - Évaluation des besoins
 - Santé publique
 - Qualité, accès, évaluation des soins de santé
-

Président : Professeur PELACCIA Thierry

Assesseurs :

- Docteur FABACHER Thibaut, AHU
 - Docteure CHAMBE Juliette, MCU
 - Docteure LECOINTRE Lise, MCU
-

Adresse de l'auteur : 1 rue des jardins 57870 WALSCHEID, France