

UNIVERSITÉ DE LILLE
UFR3S-MÉDECINE
Année : 2025

**THÈSE POUR LE DIPLÔME D'ÉTAT
DE DOCTEUR EN MÉDECINE**

**Amplification d'un signal de pharmacovigilance à partir d'un entrepôt de
données hospitalier : application à la détection des cas d'infections
bactériennes associées à la prise d'anti-inflammatoires non stéroïdiens**

Présentée et soutenue publiquement le 14 octobre 2025 à 14 heures
au Pôle Formation

Par Julien ELICES-DIEZ

JURY

Président :

Monsieur le Professeur Philippe AMOUYEL

Assesseurs :

Monsieur le Professeur Grégoire FICHEUR
Monsieur le Docteur Aghilès HAMROUN

Directrice de thèse :

Madame le Docteur Johana BÉNÉ

Avertissement

La Faculté n'entend donner aucune approbation aux opinions émises dans les thèses : celles-ci sont propres à leurs auteurs.

Sigles, acronymes et abréviations

AINS	Anti-inflammatoire non stéroïdien
ANSM	Agence nationale de sécurité du médicament et des produits de santé
BNPV	Base nationale de pharmacovigilance
CNIL	Commission nationale de l'informatique et des libertés
CRPV	Centre régional de pharmacovigilance
DCI	Dénomination commune internationale
EDS	Entrepôt de données de santé
EIM	Effet indésirable médicamenteux
EMA	<i>European Medicines Agency</i>
LLM	<i>Large Language Model</i>
PMSI	Programme de médicalisation des systèmes d'informations
PRAC	<i>Pharmacovigilance Risk Assessment Committee</i>
RCP	Résumé des caractéristiques du produit
SIH	Système d'information hospitalier
SNIIRAM	Système national d'information inter-régimes de l'Assurance Maladie
UF	Unité fonctionnelle

Sommaire

Avertissement.....	2
Remerciements	Erreur ! Signet non défini.
Sigles, acronymes et abréviations.....	3
Sommaire	4
Introduction.....	6
1 Épidémiologie des AINS	6
1.1 Définition et classification.....	6
1.2 Principales molécules et voies d'administration	7
1.3 Données de consommation en France	7
1.4 Automédication : spécificités et limites pour la surveillance	8
1.5 Profil de sécurité	8
2 Risque infectieux associé aux AINS.....	9
2.1 Mécanismes physiopathologiques potentiels	9
2.1.1 Masquage symptomatique	9
2.1.2 Perturbation de l'immunité.....	10
2.1.3 Dissémination bactérienne	10
2.2 Recommandations actuelles	10
2.2.1 Données épidémiologiques descriptives	10
2.2.2 Mesures réglementaires	11
3 La pharmacovigilance en France	12
3.1 Principes et méthodes.....	12
3.1.1 Signalement par notification spontanée	12
3.1.2 Détection de signal en pharmacovigilance	13
3.1.3 Méthodes d'imputabilité.....	14
3.1.4 Quantification du risque : apports de la pharmaco-épidémiologie	14
3.2 Limites méthodologiques : AINS et risque infectieux	15
4 Entrepôts de données de santé hospitaliers	16
4.1 Architecture	16
4.2 Typologie des données	16
4.3 Positionnements respectifs des EDS et SNDS pour l'étude des EIM liés aux AINS	17
5 Grands modèles de langue	18
5.1 Architecture et fonctionnement.....	18
5.2 Extraction d'information à partir de textes médicaux.....	19
6 Rationnel de l'étude	19
7 Objectif de l'étude	19
Matériel et méthodes	20

1	Source de données	20
2	Design de l'étude	20
3	Population	21
3.1	Critères de sélection	21
3.2	Requête dans l'EDS et requête textuelle	21
3.2.1	Terminologies recherchées	22
3.2.2	Extraction du texte	24
4	Annotation experte : classification de référence	25
4.1	Classification des cas	25
4.2	Processus d'annotation	26
5	Annotation automatisée : stratégie de détection	27
5.1	Modèle de langue et paramètres	28
5.2	Formulation initiale des instructions	28
5.3	Plan d'optimisation finale des instructions	30
6	Description des cas éligibles : analyse exploratoire	30
7	Analyses statistiques	31
	Résultats	32
1	Population	32
1.1	Diagramme de flux	32
1.2	Description de la population retenue pour annotation	33
2	Annotation experte : classification de référence	33
3	Annotation automatisée : stratégie de détection	34
3.1	Capacité discriminante sur l'ensemble d'apprentissage	34
3.1.1	Selon la classification déclarée	34
3.1.2	Selon la classification théorique	35
3.2	Capacité discriminante sur l'ensemble de test	35
4	Description des cas éligibles : analyse exploratoire	36
	Discussion	39
1	Principaux résultats	39
2	Discussion des résultats	39
3	Discussion de la méthode	41
3.1	Forces	41
3.2	Limites	41
4	Perspectives	43
	Liste des tables	45
	Liste des figures	46
	Références	47
	Annexes	50

Introduction

L'exploitation d'un entrepôt de données de santé hospitalier peut-il faciliter la détection de cas éligibles à une analyse en pharmacovigilance ? C'est la question que nous avons explorée concernant une problématique encore mal documentée avec les anti-inflammatoires non stéroïdiens : le risque de complications infectieuses.

1 Épidémiologie des AINS

1.1 Définition et classification

Les anti-inflammatoires non stéroïdiens (AINS) constituent une classe thérapeutique hétérogène définie par opposition aux corticoïdes¹. En France, ils figurent parmi les médicaments les plus couramment utilisés pour leurs propriétés anti-inflammatoires, antalgiques, antipyrétiques. Leur mécanisme d'action commun repose sur l'inhibition de la transformation de l'acide arachidonique en prostaglandines par blocage des enzymes cyclo-oxygénases COX-1 et COX-2².

La COX-1, constitutive, assure une production basale de prostaglandines. Elle est impliquée dans le maintien de fonctions physiologiques telles que la protection de la muqueuse gastrique, la régulation hémodynamique au niveau rénal, ou encore l'agrégation plaquettaire. L'inhibition de COX-1 est responsable de la majorité des effets indésirables des AINS : la diminution de la synthèse de prostaglandines gastroprotectrices expose à un risque d'ulcération gastroduodénale. L'altération de la régulation rénale peut conduire à une insuffisance rénale fonctionnelle, en particulier chez les patients à risque.

La COX-2, inductible par la présence de cytokines pro-inflammatoires (IL-1, TNF- α , endotoxines bactériennes), est responsable de la synthèse accrue de prostaglandines lors de l'inflammation. L'inhibition de la COX-2 par les AINS supprime la production des prostaglandines pro-inflammatoires, principalement la PGE2, médiateur central de la réponse inflammatoire. L'inhibition de COX-2 explique les effets thérapeutiques recherchés des AINS : la diminution de synthèse des prostaglandines pro-inflammatoires (PGE2, PGI2) réduit la vasodilatation, l'œdème et la sensibilisation des nocicepteurs périphériques, tandis que l'effet antipyrétique résulte de l'inhibition de la synthèse de PGE2 au niveau hypothalamique, centre de la thermorégulation³.

La classification des AINS repose historiquement sur leur sélectivité pour COX-1 ou COX-2 : on distingue d'une part les inhibiteurs dits « non-sélectifs », et d'autre part les inhibiteurs « COX-2 sélectifs » appelés coxibs. Les AINS les plus couramment utilisés sont non-sélectifs. Les coxibs ont été développés pour contourner la toxicité gastro-intestinale induite par les AINS non sélectifs. Toutefois il s'est avéré que le risque thrombo-embolique était plus élevé avec ces médicaments, notamment en raison d'une absence d'effet anti-agrégant plaquettaire¹.

Les AINS commercialisés en France appartiennent à plusieurs familles chimiques aux propriétés pharmacologiques distinctes. Les dérivés arylcarboxyliques constituent le groupe

le plus important, incluant les dérivés de l'acide propionique (ibuprofène, kétoprofène, naproxène), de l'acide acétique (diclofénac, acéclofénac) et les oxicams (piroxicam, méloxicam). Les salicylés, dont l'aspirine (acide acétylsalicylique) est le principal représentant, forment une classe historique des AINS¹. Le célecoxib, le parécoxib et l'étoricoxib sont les trois coxibs encore commercialisés en France.

1.2 Principales molécules et voies d'administration

L'ibuprofène est l'AINS le plus consommé en France, et le deuxième antalgique le plus consommé après le paracétamol⁴. Le profil de sécurité favorable de l'ibuprofène aux doses antalgiques explique sa large disponibilité en automédication. Depuis janvier 2020, les spécialités à base d'ibuprofène ne sont plus en libre accès en pharmacie mais restent disponibles sans ordonnance derrière le comptoir⁵.

Le kétoprofène est le deuxième AINS le plus consommé après l'ibuprofène. Il présente une activité anti-inflammatoire supérieure à l'ibuprofène à doses équivalentes, mais se distingue par un potentiel photosensibilisant particulièrement marqué avec les formes topiques⁶.

Le diclofénac présente une puissante activité anti-inflammatoire, mais aussi un risque cardiovasculaire élevé comparable à celui des coxibs⁷. Il présente des contre-indications cardiovasculaires renforcées, incluant les antécédents d'infarctus du myocarde, d'accident vasculaire cérébral, d'artériopathie périphérique et d'insuffisance cardiaque.

Le naproxène présente le profil de sécurité cardiovasculaire le moins défavorable parmi les AINS non-sélectifs, possiblement lié à un effet antiagrégant plaquettaire prolongé⁸. Cette relative sécurité cardiovasculaire en fait un choix préférentiel chez les patients à risque, bien que le risque gastro-intestinal demeure comparable aux autres AINS non-sélectifs.

La voie orale représente la principale modalité d'administration des AINS, avec des formes à libération immédiate ou prolongée. Les spécialités à base d'aspirine, ibuprofène, kétoprofène et diclofénac sont les plus courantes. Les formulations sont diverses (comprimés nus ou enrobés, gélules, capsules molles, suspensions) et présentent une biodisponibilité de 60 à 95% selon la molécule et la galénique.

L'application topique sous forme de gels, crèmes ou emplâtres représente la deuxième modalité d'administration la plus courante. Elle optimise le rapport bénéfice/risque par concentrations locales élevées et passage systémique limité (< 5% de la dose orale équivalente). Le diclofénac 1%, le kétoprofène 2,5% et l'ibuprofène 5% constituent les principales spécialités disponibles.

Concernant la voie parentérale, l'acétylsalicylate de lysine (Aspégic®) représente la principale formulation injectable disponible en France. Elle est réservée aux situations d'urgences avec nécessité d'analgésie immédiate.

La voie rectale, alternative à la voie orale avec évitement du métabolisme de premier passage hépatique, demeure de moins en moins utilisée.

1.3 Données de consommation en France

Les AINS constituent la deuxième classe thérapeutique d'antalgiques la plus consommée en France après le paracétamol, avec 34 millions de boîtes d'ibuprofène vendues en 2022. Cette prévalence peut s'expliquer par la multiplicité des indications thérapeutiques (douleurs diverses, fièvre, inflammation), l'accès à la voie orale, la disponibilité en vente libre et à faible

coût de certains médicaments et la possibilité de remboursement en cas de prescription médicale.

L'analyse des données de remboursement de l'Assurance Maladie sur la période 2006-2015 révèle une consommation d'ibuprofène relativement stable et significativement inférieure à celle observée dans les pays voisins (11% du total des antalgiques non opiacés contre 59%, 32% et 25% en Allemagne, Espagne et Danemark, respectivement)⁴. Elle peut s'expliquer par des habitudes de consommation davantage marquées pour le paracétamol, qui représente 64% du marché total des antalgiques en France, et par des politiques de remboursement variables selon les systèmes de santé. La période 2005-2011 avait connu une croissance soutenue, notamment liée au retrait du dextropropoxyphène qui a entraîné un report de prescription vers d'autres antalgiques. Puis une diminution transitoire en 2020 liée à la pandémie de COVID-19 et aux recommandations restrictives concernant l'utilisation des AINS dans ce contexte⁹.

La consommation d'AINS présente des variations importantes selon les caractéristiques démographiques et socio-économiques. Les femmes en sont les principales consommatrices, notamment en raison de symptomatologies spécifiques telles que les dysménorrhées et les céphalées menstruelles¹⁰. Les adultes jeunes (20-40 ans) représentent la tranche d'âge la plus consommatrice, avec une utilisation privilégiée pour les douleurs musculo-squelettiques liées aux activités sportives et professionnelles. Les patients souffrant de pathologies rhumatismales chroniques constituent une population particulièrement exposée, avec des durées de traitement prolongées augmentant le risque d'effets indésirables.

1.4 Automédication : spécificités et limites pour la surveillance

L'automédication représente une part substantielle de la consommation d'AINS en France mais reste difficile à quantifier. Dans une étude visant à caractériser les cas d'infections graves sous AINS rapportés entre 2011 et 2013, l'automédication était présente dans 61% des cas¹¹. Une enquête menée dans la Somme en 2018 révèle que 67% des patients interrogés pratiquaient l'automédication, principalement pour traiter des céphalées (68,6%), en utilisant pour la plupart des médicaments anciennement prescrits ou par achat direct avec conseil du pharmacien¹².

Certains AINS, notamment l'ibuprofène et le kétoprofène, sont disponibles sans ordonnance et ainsi non remboursés par l'Assurance Maladie dans ce contexte. Cette accessibilité génère une limite méthodologique majeure pour la surveillance épidémiologique : le Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM), une des composantes principales du Système National des Données de Santé (SNDS), collecte l'information relative aux médicaments délivrés, uniquement pour les prescriptions remboursées¹³. Cette absence de traçabilité des ventes sans ordonnance dans les bases médico-administratives créent un angle mort dans l'évaluation exhaustive du risque d'effet indésirable médicamenteux associé aux AINS en population.

1.5 Profil de sécurité

Un effet indésirable médicamenteux (EIM) se définit comme toute réaction nocive et non recherchée survenant lors de l'utilisation d'un médicament aux posologies normales ou non, pour laquelle un lien avec le médicament est suspecté ou établi¹⁴. Il doit être distingué de

l'événement indésirable, qui désigne toute manifestation survenant chez un patient traité sans préjuger d'un lien avec le médicament. Le profil de sécurité des AINS est bien documenté depuis leur commercialisation dans les années 1960, avec plusieurs classes d'EIM identifiées.

Selon les Résumés des Caractéristiques du Produit (RCP), les EIM les plus fréquemment décrits avec les AINS sont de nature gastro-intestinale, dose-dépendants, incluant des manifestations allant de la dyspepsie aux complications potentiellement fatales (ulcères, perforations, hémorragies)¹⁵. La toxicité cardiovasculaire, particulièrement documentée depuis le retrait du rofécoxib en 2004, se manifeste par une augmentation du risque d'infarctus du myocarde, notamment avec les coxibs et le diclofénac, et surtout à doses élevées et lors de traitements prolongés. La toxicité rénale, également dose-dépendante, peut conduire à une insuffisance rénale fonctionnelle chez les patients à risque (déshydratation, association aux diurétiques, sujets âgés). Les réactions d'hypersensibilité, classées comme peu fréquentes à très rares, peuvent néanmoins inclure des manifestations graves telles que des syndromes de Lyell et de Stevens-Johnson.

Plus récemment, les RCP ont intégré au profil de sécurité des AINS un risque de complications infectieuses graves, notamment cutanées et des tissus mous en cas de varicelle, ainsi qu'un rôle potentiellement aggravant en cas de pneumonie communautaire d'origine bactérienne¹⁵. Cette problématique, moins documentée que les EIM classiques, a justifié une analyse plus approfondie et soulevé la question d'une réévaluation de la balance bénéfice-risque des AINS en contexte infectieux¹⁶.

2 Risque infectieux associé aux AINS

2.1 Mécanismes physiopathologiques potentiels

Les mécanismes par lesquels les AINS favoriseraient les complications infectieuses dépassent le simple masquage des symptômes : interaction entre inhibition COX, perturbation de l'immunité et facilitation de la dissémination bactérienne notamment streptococcique.

2.1.1 Masquage symptomatique

L'effet antipyétique des AINS lié à l'inhibition de la PGE2 hypothalamique peut entraîner un retard dans le diagnostic et la prise en charge thérapeutique de complications potentiellement graves, leur manifestation pouvant survenir 2 à 3 jours après la prise^{3,16}.

L'effet antalgique des AINS peut également être défavorable. Par exemple, l'atténuation de la douleur exquise initiale dans la fasciite nécrosante, souvent disproportionnée par rapport aux signes cliniques, peut retarder la prise en charge de l'urgence vitale¹⁷.

Ainsi, les AINS peuvent être à l'origine d'une dissociation entre la sévérité réelle de l'infection et sa présentation clinique (dissociation clinico-biologique) : alors que l'infection progresse, les signes inflammatoires locaux (rougeur, chaleur, œdème) se retrouvent atténués par l'inhibition des prostaglandines vasodilatatrices². Cette dissociation est notamment marquée dans les infections streptococciques, où la thrombose microvasculaire responsable de la nécrose tissulaire progresse silencieusement, tandis que les marqueurs biologiques d'inflammation (CRP, procalcitonine) demeurent élevés¹⁷.

Un risque majoré d'évolution défavorable malgré la co-prescription d'une antibiothérapie appropriée suggère l'existence d'autres mécanismes physiopathologiques¹⁶.

2.1.2 Perturbation de l'immunité

La PGE2, bien que pro-inflammatoire, exerce des fonctions immunomodulatrices. Elle favorise initialement l'inflammation locale par vasodilatation et l'augmentation de la perméabilité vasculaire, facilitant le recrutement des neutrophiles et macrophages au site infectieux. L'inhibition de la PGE2 par les AINS entraîne ainsi une diminution des fonctions phagocytaires et bactéricides¹⁸.

Par ailleurs, la PGE2 influence la différenciation des lymphocytes T CD4+ naïfs et la production d'IL-12 par les cellules dendritiques. Son inhibition favorise les réponses Th2 et Th17 au détriment de la réponse Th1, nécessaire pour élimination des pathogènes intracellulaires notamment streptococciques¹⁹⁻²¹.

2.1.3 Dissémination bactérienne

Les lésions tissulaires induisent la production de vimentine, protéine du cytosquelette servant de médiateur à l'adhésion locale des streptocoques du groupe A²². Les AINS pourraient amplifier l'expression de la vimentine et participer directement à la dissémination bactérienne, pouvant expliquer la prédominance streptococcique dans les complications : 62% des infections bactériennes graves sous ibuprofène impliquent des streptocoques, principalement *S. pyogenes* (50%) et *S. pneumoniae* (28%)¹⁶. L'expérimentation animale montre que l'administration d'ibuprofène augmente la sévérité des infections nécrosantes à *S. pyogenes* et diminue l'efficacité de l'antibiothérapie associée.

2.2 Recommandations actuelles

2.2.1 Données épidémiologiques descriptives

Suite aux signalements récurrents de complications infectieuses graves avec AINS utilisés dans le traitement symptomatique de la fièvre ou la douleur, l'ANSM a commandé en juin 2018 aux Centres régionaux de pharmacovigilance (CRPV) de Tours et Marseille une enquête nationale de pharmacovigilance portant spécifiquement sur l'ibuprofène et le kétoprofène, les deux AINS les plus utilisés dans ces indications²³. Cette enquête s'inscrivait dans la continuité d'un premier travail réalisé en 2002 sur le risque d'infections de la peau et des tissus mous associé aux AINS en pédiatrie en contexte de varicelle, qui avait déjà conduit à la modification du RCP en 2004.

Sur l'ensemble des cas rapportés dans la Base nationale de pharmacovigilance (BNPV) entre 2000 et 2018 (n=844), 337 cas de complications infectieuses associés à l'ibuprofène et 49 cas associés au kétoprofène ont été retenus, après sélection des cas les plus graves chez des enfants ou des adultes sans facteur de risque ni comorbidités. Ces cas comprenaient des infections sévères de la peau et des tissus mous de type cellulites, dermohypodermites et fasciites nécrosantes (40,9%), des infections pleuro-pulmonaires de type empyèmes et pneumonies compliquées (32,1%), des complications neurologiques et ORL (16,1%) ainsi que des sepsis (11,4%).

Les complications pouvaient survenir après de très courtes durées de traitement (2 à 3 jours), y compris lorsque la prise d'AINS était associée à une antibiothérapie. Les infections

impliquaient principalement des pathogènes de type streptocoque ou pneumocoque. Au-delà du simple masquage symptomatique, les auteurs évoquent l'effet amplificateur intrinsèque des AINS sur la dissémination bactérienne via l'expression de vimentine comme mécanisme pouvant expliquer au moins partiellement l'évolution défavorable malgré une antibiothérapie associée.

Une actualisation du rapport en 2024 a été demandée suite à la déclaration de 4 cas pédiatriques dont 3 décès liés à des infections à streptocoque A avec prise d'ibuprofène dans les 48 heures précédentes¹⁶. Celle-ci souligne une augmentation de la sévérité et du nombre de cas graves d'infections streptococciques associées à l'ibuprofène et au kétoprofène malgré un contexte de diminution des ventes de ces médicaments. Sur une période de 5 ans, 21% des effets indésirables graves déclarés avec l'ibuprofène concernaient une infection ou une aggravation d'infection bactérienne, contre 8% pour le kétoprofène.

Ces données contrastent avec la persistance de pratiques à risque en médecine de ville. L'analyse des données de la Caisse Primaire d'Assurance Maladie (CPAM) sur l'année 2019 montre que 13,25 % des prescriptions d'antibiotiques en soins primaires comportent une co-prescription d'AINS. Les médecins généralistes représentent la source principale de ces co-prescriptions, étant responsables de 74,6 % des prescriptions d'antibiotiques de médecine de ville, devant les médecins spécialistes (12,6 %), les chirurgiens-dentistes (12,5 %).

2.2.2 Mesures réglementaires

2.2.2.1 *Recommandations nationales*

Selon le RCP français de l'ibuprofène, les règles de prescription imposent : une dose minimale efficace, une durée de traitement sans avis médical limitée à 5 jours en cas de douleur et 3 jours en cas de fièvre, l'arrêt dès disparition des symptômes et une consultation immédiate en cas d'aggravation¹⁵.

Le point d'information de l'ANSM du 18 avril 2019, actualisé en 2023, établissait la recommandation de privilégier systématiquement le paracétamol en cas de douleur ou fièvre dans tout contexte infectieux (angine, rhinopharyngite, otite, infection pulmonaire, dentaire, cutanée, varicelle)²⁴. La lettre aux professionnels de santé de juillet 2004, qui établissait la contre-indication des AINS en cas de varicelle, a été confirmée et étendue en 2019²⁵.

En janvier 2020, l'ANSM ordonnait le retrait des médicaments contenant de l'ibuprofène, du paracétamol et de l'aspirine du libre accès en pharmacie^{5,26}. Cette mesure préventive unique en Europe visait à renforcer un conseil pharmaceutique obligatoire avant toute délivrance et à contrôler davantage le phénomène d'automédication.

2.2.2.2 *Recommandations européennes*

Suite à l'enquête menée par l'ANSM, l'Agence européenne des médicaments (EMA) a initié en mai 2019 une procédure d'évaluation portant sur l'ibuprofène et le kétoprofène via son Comité d'évaluation des risques en pharmacovigilance (PRAC)²⁷. En avril 2020, le PRAC a validé les conclusions françaises en reconnaissant que la prise d'ibuprofène ou de kétoprofène pouvait favoriser un masquage symptomatique de la fièvre et de la douleur lors de certaines infections et ainsi occasionner un retard de diagnostic et de prise en charge.

Une harmonisation des recommandations et notices patient a été imposée pour tous les médicaments contenant de l'ibuprofène ou du kétoprofène. Ces modifications incluent l'ajout

d'une section spécifique intitulée « Dissimulation des symptômes d'une infection sous-jacente » dans la rubrique 4.4 des RCP, qui mentionne explicitement les pneumonies communautaires bactériennes et les complications bactériennes de la varicelle comme situations à risque documentées¹⁵.

2.2.2.3 Divergences nationales et européennes

Cette actualisation au niveau européen n'a cependant pas mené au retrait des médicaments du libre accès en pharmacie, comme ce fut le cas en France. Tandis que l'ANSM adopte une démarche préventive en s'appuyant sur l'expertise des CRPV, qui privilégie la détection précoce des signaux au prix de l'incertitude, l'EMA adopte une approche plus conservatrice et exige un processus d'évaluation nécessitant une validation selon des critères prédefinis²⁸. Les divergences entre agences nationales et européennes portent notamment sur le nombre de cas documentés jugé insuffisant selon les standards européens, une hétérogénéité dans les définitions et critères diagnostiques, et le manque de données issues d'études épidémiologiques à fort niveau de preuve.

L'exemple de la pandémie de COVID-19 a pu illustrer ces différentes approches : alors qu'en mars 2020 l'ANSM déconseillait précocement, par mesure de précaution, l'utilisation des AINS chez les patients atteints de COVID-19, l'EMA tenait une position plus nuancée en déclarant qu'il n'existe aucune preuve scientifique d'un lien entre utilisation d'ibuprofène et aggravation de la maladie²⁹.

3 La pharmacovigilance en France

3.1 Principes et méthodes

Le système national français de pharmacovigilance créé en 1973 est un dispositif de surveillance et de prévention des effets indésirables médicamenteux^{30,31}. En France, l'activité de pharmacovigilance est à la fois centralisée au niveau de l'Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM) et décentralisée grâce à l'activité des 30 CRPV répartis sur le territoire.

Les CRPV recueillent et analysent les déclarations relatives aux effets indésirables médicamenteux. Ces déclarations sont effectuées de manière spontanée par les professionnels de santé et les patients, quelle que soit la gravité et la nature des effets et qu'ils soient mentionnés ou non dans les monographies des médicaments. Après analyse par un pharmacologue sur 3 principaux critères d'imputabilité (chronologique, sémiologique et bibliographique), ces déclarations sont enregistrées dans la BNPV de manière anonyme pour le patient comme pour le déclarant³². Cette base de données existe depuis 1984 et contient à ce jour un peu plus de 1 540 000 cas.

3.1.1 Signalement par notification spontanée

La notification spontanée des EIM est le dispositif central de déclaration en pharmacovigilance³³. Une déclaration valide nécessite quatre éléments : un déclarant identifiable, un patient identifiable, un médicament suspect et la description de l'effet observé. Le circuit de traitement suit une procédure standardisée : validation de la complétude du cas,

évaluation clinique par le CRPV, attribution d'un score d'imputabilité, transmission à l'ANSM puis vers EudraVigilance dans un délai de 15 jours pour les EIM graves³⁴.

Ce dispositif a pour avantage de permettre une surveillance continue de tous les médicaments commercialisés et la détection précoce de signaux émergents, avec un faible coût et une simplicité d'utilisation. Les médecins, pharmaciens, chirurgiens-dentistes, sages-femmes et infirmiers sont soumis à une obligation légale de déclaration des effets indésirables, quels qu'ils soient³⁵. L'enrichissement par les déclarations patients, particulièrement visible lors de la surveillance des vaccins COVID-19, apporte une dimension complémentaire³⁶.

La sous-notification constitue cependant une limite structurelle majeure, avec un taux estimé à 95 % pour l'ensemble des EIM et 80% pour les EIM graves³³. Bien qu'ils diagnostiquent la majorité des EIM, les médecins généralistes étaient à l'origine de seulement 7% des déclarations de pharmacovigilance réalisées en 2014³⁷. Plusieurs facteurs expliquant ce phénomène ont déjà été décrits, parmi lesquels une conviction que les effets graves sont déjà bien documentés et que seuls eux devraient être signalés, une difficulté à percevoir le lien causal ou encore par manque de temps ou d'intérêt³⁸.

La qualité hétérogène de l'information présente dans ces notifications constitue une autre limite. Sur les 3000 déclarations de pharmacovigilance réalisées par les médecins généralistes en 2015, seulement 16,4% étaient considérées complètes³⁹. Les déclarations complètes étaient significativement associées aux cas graves et à l'âge du patient (personne âgée et enfant).

3.1.2 Détection de signal en pharmacovigilance

Un signal est défini selon l'EMA comme toute information mentionnant un risque dont le caractère nouveau ou inhabituel justifie une investigation²⁸. Cette définition englobe les aspects qualitatifs (nature de l'effet, mécanisme) et quantitatifs (fréquence, force de l'association), et peut émerger de sources multiples : notifications spontanées, littérature scientifique, études pharmaco-épidémiologiques, bases de données administratives.

Le processus de validation d'un signal en pharmacovigilance peut être résumé comme suit :

1. Détection initiale : un signal émerge de sources diverses transmises par les professionnels de santé ou les patients aux CRPV, constituant une première alerte basée sur l'observation de cas individuels.
2. Caractérisation clinique : analyse approfondie des cas rapportés pour identifier le profil démographique et clinique des patients concernés, établir la chronologie entre l'exposition médicamenteuse et la survenue de l'effet, et documenter l'évolution après arrêt du traitement.
3. Confirmation épidémiologique : mise en œuvre d'études observationnelles sur de larges cohortes permettant de quantifier l'augmentation du risque dans la population exposée, comparativement à la population non exposée. Il est précisé que le niveau de preuve de ce type d'étude est supérieur à celui-ci obtenu dans le cadre des détections ou caractérisations décrites en 1. et 2.

Les approches traditionnelles quantitatives de détection de signal utilisent l'analyse de disproportionnalité⁴⁰. Cette méthode compare la fréquence de notification d'un couple médicament-EIM à celle attendue dans la base de pharmacovigilance, permettant d'identifier des associations statistiquement disproportionnées. Deux paradigmes coexistent :

- Les méthodes fréquentistes utilisent principalement le Proportional Reporting Ratio (PRR) et le Reporting Odds Ratio (ROR). Un signal est considéré comme significatif lorsque le PRR est supérieur ou égal à 2, accompagné d'un χ^2 supérieur ou égal à 4 et d'au moins 3 cas notifiés. Le ROR, dont le calcul s'apparente à celui d'un odds ratio classique, utilise des seuils similaires avec une borne inférieure de l'intervalle de confiance à 95% supérieur ou égal à 1.
- Les méthodes bayésiennes, notamment l'Information Component (IC) développé par l'OMS, appliquent des techniques de « shrinkage bayésien » afin de limiter l'influence des événements rares. Un signal est détecté lorsque la borne inférieure de l'intervalle de crédibilité à 95% (IC025) est positive⁴¹.

L'application de ces méthodes dans la base de pharmacovigilance génèrent des signaux statistiques soumis ensuite à une expertise clinique approfondie. Ces analyses présentent toutefois des limites importantes : elles génèrent des hypothèses plutôt que des preuves de causalité, sont sensibles aux biais de notification et ne permettent pas d'estimer l'incidence réelle des EIM. Les recommandations READUS-PV, publiées en 2024, établissent pour la première fois des standards pour décrire les analyses de disproportionnalité dans les études, en préconisant notamment une taille minimale de 500 notifications pour les bases nationales afin de limiter les fausses associations^{40,42}.

3.1.3 Méthodes d'imputabilité

La méthode française, développée par Dangoumau en 1978, actualisée par Bégaud en 1985 puis par Arimone en 2011, constitue une spécificité nationale³². L'analyse chronologique (score C) évalue trois critères : délai de survenue, évolution à l'arrêt (« dechallenge ») et à la réadministration (« rechallenge »). L'évaluation sémiologique (score S) analyse la plausibilité biologique, les facteurs favorisants et l'exclusion des diagnostics différentiels. La combinaison des scores génère 7 niveaux d'imputabilité allant de « exclue » à « très vraisemblable ».

La méthode d'imputabilité étasunienne de Naranjo est basée sur un questionnaire qui renseigne sur la séquence chronologique, l'évolution au « dechallenge » et au « rechallenge », les diagnostics différentiels, l'effet à une prise de placebo, la relation dose-effet, l'expérience antérieure du patient avec le médicament et la confirmation par des éléments objectifs⁴³.

Il existe des divergences entre les différentes méthodes d'imputabilité. Une analyse de 188 cas impliquant le voriconazole et notifiés entre 2002 et 2005 retrouve une concordance quasi-nulle entre la méthode française et celle de Naranjo ($\kappa = 0,037$)⁴⁴. La méthode française privilégie la sensibilité pour détecter les effets nouveaux, au prix d'une spécificité moindre générant de nombreux scores « douteux ».

3.1.4 Quantification du risque : apports de la pharmaco-épidémiologie

La pharmacovigilance et la pharmaco-épidémiologie constituent deux approches complémentaires dans la surveillance de la sécurité du médicament. Elles se distinguent fondamentalement par leur démarche méthodologique⁴⁵.

La pharmacovigilance suit une approche exploratoire et individuelle, basée sur l'analyse de cas et la formulation d'hypothèses à partir de signaux émergents, sans nécessiter de question de recherche prédéfinie. Cette démarche permet la détection précoce d'effets indésirables rares ou inattendus, comme l'illustre l'identification des myocardites post-vaccinales COVID-

19, où quelques dizaines de cas bien documentés ont suffi à générer un signal avant toute confirmation épidémiologique³⁶.

À l'inverse, la pharmaco-épidémiologie suit une approche populationnelle, en testant des hypothèses prédéfinies par des études observationnelles permettant de quantifier le risque et d'établir des associations causales robustes⁴⁵. Classiquement, trois grands types d'études sont mobilisés :

- Les études de cohorte prospectives comparent l'incidence d'un événement indésirable entre des populations exposées et non exposées au médicament, permettant le calcul direct du risque relatif et du risque absolu.
- Les études cas-témoins, particulièrement adaptées aux événements rares, comparent rétrospectivement l'exposition médicamenteuse entre des patients présentant l'effet indésirable (cas) et des sujets indemnes (témoins), fournissant un odds ratio qui approxime le risque relatif lorsque l'incidence est faible.
- Les études transversales mesurent la prévalence de l'utilisation médicamenteuse et des effets indésirables à un instant donné dans une population.

Cette méthodologie permet d'établir l'existence d'une association statistique, de quantifier précisément l'excès de risque, d'ajuster sur les facteurs de confusion par analyses multivariées et d'évaluer la relation dose-effet. Malgré des coûts plus élevés en temps et en effectifs, elle offre un support indispensable aux décisions réglementaires notamment au niveau européen³⁴.

3.2 Limites méthodologiques : AINS et risque infectieux

En raison des limites précédemment décrites, inhérentes au système de pharmacovigilance, le nombre réel de cas d'infections bactériennes associées à la prise d'AINS est vraisemblablement sous-estimé. Au niveau national, moins d'un cas est notifié en moyenne dans la BNPV sur une période de 5 ans. La disponibilité déjà évoquée de certains AINS en automédication, ainsi que le manque de sensibilisation des professionnels de santé à la problématique participent au fait que la consommation réelle d'AINS est insuffisamment mentionnée dans les comptes rendus médicaux, contrairement par exemple à la prise d'antibiotiques.

D'autre part, l'attribution de la complication infectieuse à l'AINS plutôt qu'à l'évolution naturelle de la maladie se heurte à un biais dit de temporalité inverse (ou biais protopathique), pouvant se manifester lorsqu'un AINS est administré pour soulager les symptômes précurseurs d'une infection déjà en cours mais encore non diagnostiquée. L'imputabilité se complique davantage en cas de traitements concomitants.

Face à ces limites compromettant une évaluation fiable et exhaustive du risque infectieux associé aux AINS, il devient nécessaire de recourir à d'autres approches méthodologiques.

4 Entrepôts de données de santé hospitaliers

4.1 Architecture

Un entrepôt de données de santé hospitalier (EDS) est une infrastructure centralisée dont la finalité est de collecter, structurer et mettre à disposition les données médico-administratives générées lors du parcours de soins des patients au sein d'un établissement hospitalier⁴⁶. Ces données, initialement collectées à des fins de soins, peuvent être réutilisées à des fins de pilotage de l'activité hospitalière, de recherche biomédicale et d'amélioration de la qualité des soins. Le développement des EDS en France s'est considérablement renforcé à partir de la fin des années 2010. Au 1er janvier 2025, on dénombre une centaine d'EDS ayant obtenu une autorisation de la Commission nationale de l'informatique et des libertés (CNIL), dont 17 constitués au sein d'un centre hospitalo-universitaire (CHU)⁴⁷.

L'architecture technique des EDS repose sur plusieurs modèles standardisés internationaux permettant l'interopérabilité et la réutilisation des données de santé. Les deux principaux protocoles d'interopérabilité utilisés en France sont *Informatics for Integrating Biology and the Bedside* (i2b2) et *Observational Medical Outcomes Partnership* (OMOP), l'un facilitant l'intégration de données hétérogènes issues des différents systèmes d'information hospitaliers (SIH), l'autre s'imposant comme standard pour la recherche clinique observationnelle^{46,48}.

4.2 Typologie des données

L'intégration multi-sources des données dans l'EDS, incluant à la fois des données structurées et non structurées, garantit une couverture exhaustive et en temps réel de l'activité hospitalière. Les données dites « structurées » correspondent à un ensemble d'informations codifiées selon des terminologies et formats standardisés. Il s'agit principalement :

- Données du PMSI :
 - Codage des diagnostics, utilisant la Classification Internationale des Maladies 10e révision (CIM-10)
 - Codage des actes médicaux, utilisant la Classification Commune des Actes Médicaux (CCAM)
 - Données relatives aux séjours : résumés d'unité médicale (RUM), regroupés en résumé de sortie standardisé (RSS)
- Résultats biologiques et anatomo-pathologiques
- Historiques des prescriptions et administrations médicamenteuses
- Rapports d'interventions et fiches d'évaluation diverses

Ces données rapidement exploitables offrent une standardisation optimale au prix d'une granularité clinique limitée.

Les données dites « non structurées » sont issues du dossier patient informatisé (DPI), qui constitue environ 80% du volume de données hospitalières⁴⁹. Le DPI centralise l'ensemble des informations médicales, paramédicales et administratives produites lors du parcours de soins du patient (comptes rendus médicaux, lettres de sortie, historiques de soins infirmiers...), remplaçant progressivement le dossier papier. En pratique, il intègre à la fois des champs de saisie structurée (prescriptions, constantes) et des champs de saisie libre permettant aux rédacteurs de consigner leurs observations détaillées et développer un narratif. Ce type

d'information est qualifié de « texte libre ».

Contrairement aux données structurées limitées à des champs prédéfinis, les données textuelles permettent de capturer l'information médicale dans sa complexité. Elles ont cependant des degrés variables d'information et de qualité rédactionnelle qui découlent de leur utilisation en pratique :

- Les observations médicales et notes de soins infirmiers : plutôt informel et soumis à variabilité individuelle car réservé à l'équipe de soin, mais potentiellement riche en détails cliniques, notamment paramédicaux.
- Les comptes rendus (hospitalisations, opérations, imagerie) : intermédiaire
- Les lettres de sortie ou de liaison, courriers de consultation : plus formels et homogènes dans la rédaction mais aussi plus synthétiques.

Le caractère « structuré » ou « non structuré » de la donnée est cependant à différencier de la qualité rédactionnelle : une observation médicale (soit du texte libre) peut être rédigée de façon structurée. Les données non structurées peuvent être multimodales, intégrant à la fois du texte libre et des éléments visuels (radiographies, scanners, photographies, tracés numérisés).

4.3 Positionnements respectifs des EDS et SNDS pour l'étude des EIM liés aux AINS

L'information relative à la consommation sur le mode de l'automédication paraît essentielle pour documenter de manière exhaustive le risque lié à l'exposition aux AINS, largement disponibles sans ordonnance. Comme décrit précédemment, les données de remboursement collectées dans le SNDS ne sont pas en mesure de capturer cette information ; il en va même pour les données structurées collectées par les SIH, dont le champ est restreint à la période du séjour hospitalier.

L'exploitation des données textuelles non structurées, via les comptes rendus et observations médicales, semble adaptée pour palier à cette problématique. Le projet HOPIPRAC du CHU de Brest, via l'analyse du DPI sur la période 2015-2018, a permis de retrouver 98% des cas enregistrés dans la BNPV, d'identifier des signaux émergents tels que 34 cas d'effets indésirables sous nivolumab ou encore des associations médicament-effet non déclarées dans 56% des cas de toxidermies bulleuses analysées⁵⁰.

La flexibilité de l'EDS procure certains avantages opérationnels. Contrairement au SNDS dont la structure est fixée au niveau national, un EDS peut rapidement intégrer de nouvelles sources de données et s'adapter aux besoins spécifiques de recherche. De plus la disponibilité en temps réel ou quasi-temps réel des données contraste avec les délais de consolidation du SNDS, qui peuvent varier de 3 à 12 mois⁴⁶. Cette réactivité facilite la réalisation d'études nécessitant la mise en place de systèmes d'alerte précoce.

Néanmoins, l'EDS hospitalier ne permet pas le suivi longitudinal des patients exposés à un AINS pour identifier les complications infectieuses ultérieures, contrairement à ce que permet une cohorte rétrospective construite dans le SNDS. En effet, il est courant qu'un même patient soit suivi ensuite en ville ou dans un autre hôpital. Ainsi, il existe un biais d'observation et une censure informative quasi-systématique si l'on considère un suivi débutant au cours d'une hospitalisation. Cette discontinuité justifie la réalisation d'une étude s'appuyant sur les données d'un seul temps hospitalier, par exemple en partant des cas d'infections documentées

lors d'une hospitalisation pour rechercher rétrospectivement, au sein des documents de ce même séjour, une exposition récente à un AINS.

L'exhaustivité permise par l'EDS s'accompagne également d'enjeux méthodologiques : le volume de données produit quotidiennement et leur nature hétérogène rendent la recherche d'information manuelle difficilement réalisable, voire impossible en pratique. Ce constat impose le recours à des méthodes de traitement automatique du langage naturel pour faciliter la détection de cas d'intérêt.

5 Grands modèles de langue

5.1 Architecture et fonctionnement

Les grands modèles de langue (*Large Language Model*, LLM), dont le développement s'est accéléré depuis 2020 avec la sortie de GPT-3 et la démocratisation de ChatGPT fin 2022, constituent une évolution dans le traitement automatique du langage naturel⁵¹. Contrairement aux approches traditionnelles basées sur des règles prédéfinies, les LLM apprennent les structures linguistiques et les relations sémantiques à partir d'entraînements sur des corpus de texte massifs. Cette capacité d'apprentissage permet l'identification de patterns complexes dans des textes spécialisés, même face à la variabilité terminologique et syntaxique inhérente à la donnée non structurée.

Les LLM reposent sur une architecture à réseaux de neurones de type « *transformer* », modèles complexes qui analysent et transforment progressivement l'information textuelle^{52,53}. Ils peuvent être catégorisés selon leur mode d'accès : les modèles commerciaux (GPT-4, Claude) sont soumis à des restrictions d'usage mais offrent généralement des performances optimales, tandis que les modèles open source (Llama, Mistral) permettent une implémentation locale garantissant la confidentialité des données et un meilleur contrôle des paramètres⁵⁴. Ils peuvent également se distinguer selon leur modalité : les modèles unimodaux traitent exclusivement du texte, tandis que les modèles multimodaux sont capables d'intégrer simultanément plusieurs types de données (texte, images, audio, vidéo).

L'interaction avec un LLM s'effectue via un « *prompt* », anglicisme qui désigne un ensemble structuré d'instructions guidant le LLM dans son traitement de l'information⁵⁵. Dans une tâche de classification, il doit être clairement défini l'objectif à suivre, les critères de décision à appliquer et le format de réponse attendu. La précision de ces instructions conditionne directement les performances du LLM dans la réalisation de la tâche demandée. Le « *prompt engineering* » désigne le processus itératif d'optimisation des instructions visant à maximiser les performances du LLM dans une tâche précise.

Un LLM traite le texte par unités élémentaires appelées « *tokens* », qui correspondent généralement à un fragment de mot, un mot entier ou un symbole de ponctuation⁵⁶. La tokenisation impacte notamment la taille de contexte disponible et les coûts d'utilisation des modèles commerciaux. Un LLM génère du texte par prédiction successive de tokens selon leur probabilité de distribution sur l'ensemble du vocabulaire, à l'image d'un compléteur de texte⁵³. Ce mécanisme probabiliste implique la production de réponses potentiellement variables pour un même prompt (comportement dit « *stochastique* », par opposition à

déterministe). Cette variabilité est modulable via un paramétrage du modèle ou des techniques de *prompt engineering*.

5.2 Extraction d'information à partir de textes médicaux

L'extraction d'information complexe à partir de données médicales non structurées rentre dans le champ d'application des LLM. Au-delà de la simple identification d'entités (médicaments, pathologies, symptômes), ces modèles sont capables d'analyser le contexte clinique dans lequel ces éléments apparaissent. Des modèles spécialisés dans le français médical comme DrBERT et CamemBERT-bio ont été développés pour pallier aux limites des modèles généralistes^{57,58}.

Ces modèles peuvent entre autres distinguer efficacement un épisode actuel d'un antécédent, ou une hypothèse diagnostique d'un diagnostic confirmé. Cette compréhension contextuelle s'étend aux relations temporelles et aux nuances linguistiques comme les négations ou incertitudes, souvent présentes dans les textes médicaux. Les LLM sont ainsi capables de transformer des volumes massifs de données textuelles inexploitables manuellement en informations structurées analysables à grande échelle.

6 Rationnel de l'étude

D'une part, les AINS constituent une classe médicamenteuse largement consommée en population générale, tant sur le mode de la prescription médicale qu'en automédication. D'autre part, l'EDS est en mesure de capturer de façon exhaustive les séjours hospitaliers en lien avec un épisode infectieux en cours, tout en conservant dans les textes médicaux l'information relative à une potentielle consommation d'AINS au préalable. Le recours à un LLM pour le traitement de ces données textuelles massivement produites pourrait faciliter la détection de cas de complications infectieuses potentiellement imputables à une prise d'AINS.

7 Objectif de l'étude

L'objectif principal de l'étude était de développer et évaluer une méthode de détection de cas éligibles à une analyse d'imputabilité en pharmacovigilance pour la recherche d'EIM à type d'infection bactérienne en lien avec une prise d'AINS. Cette méthode de détection s'appuyait sur l'identification de cas d'infections bactériennes concomitantes de la prise d'AINS, à partir d'un EDS hospitalier.

Les objectifs secondaires étaient de décrire les cas classés éligibles, et de les croiser avec les déclarations de pharmacovigilance réalisées au CRPV du CHU de Lille sur la même période.

Matériel et méthodes

1 Source de données

L'étude exploitait les données de l'EDS du CHU de Lille (INCLUDE), dont le déploiement a été autorisé par la CNIL en 2019⁵⁹. Les vigilances font partie des typologies d'études ayant été explicitement prévues lors de la demande formulée à la CNIL. L'EDS INCLUDE repose sur une architecture de base de données relationnelle permettant le stockage structuré et l'interrogation complexe de données médico-administratives massives. L'architecture relationnelle organise les données en de multiples tables interconnectées, autorisant la formulation de requêtes croisées via le langage SQL (*Structured Query Language*).

L'EDS centralise l'ensemble des données relatives aux séjours hospitaliers réalisés depuis 2016 dans le secteur Médecine-Chirurgie-Obstrétiqe (MCO), et produites par les différents SIH de l'hôpital, en particulier Sillage (SIH principal gérant les données médico-administratives), MOLIS (résultats biologiques et anatomo-pathologiques), QDOC (résultats d'imagerie), Resurgence (urgences), DIANE et ICCA (anesthésie-réanimation). En pratique, les documents produits par la plupart des SIH sont inclus dans Sillage avant leur intégration dans l'EDS. Chaque document stocké dans l'EDS est rattaché à un séjour, qui est lui-même rattaché au patient.

Seuls les documents produits et signés numériquement sont intégrés à l'EDS. Depuis 2024, cette intégration est réalisée quotidiennement et quasi-instantanément (les documents signés la veille sont intégrés dans l'EDS le lendemain). L'interrogation de l'EDS par le biais de requêtes SQL permet l'extraction ciblée et standardisée de corpus de documents selon des critères cliniques, temporels et administratifs bien définis.

Dans le cadre de cette étude, nous nous sommes restreints à l'extraction et l'analyse de données non structurées exclusivement textuelles.

2 Design de l'étude

Dans cette étude monocentrique, afin d'évaluer la capacité discriminante de la stratégie de détection proposée, deux annotations étaient réalisées en aveugle l'une de l'autre :

- Une annotation dite « experte », utilisée comme référence (*gold standard*).
- Une annotation dite « automatisée », réalisée par un modèle de type LLM.

L'ensemble de la méthode employée pouvait être résumée et subdivisée en trois phases successives :

1. Constitution d'un corpus de documents à annoter

- Requête en 2 temps :
 - Critères d'inclusion et de non-inclusion : extraction par une requête dans l'EDS
 - Critère d'exclusion : affinement par une requête textuelle sur les documents inclus (AINS, infections et agents pathogènes)

2. **Annotation experte des documents : classification de référence**
 - Classification binaire réalisée par 2 experts : cas « éligibles » vs « inéligibles »
 - Description des cas classés éligibles (analyse exploratoire)
3. **Annotation automatisée des documents : stratégie de détection**
 - Classification binaire réalisée par un LLM
 - *Prompt engineering* et évaluation des capacités discriminantes

3 Population

3.1 Critères de sélection

Critères de non-inclusion :

- Tout document massivement produit en routine et dépourvu de donnée non structurée, ou n'apportant pas d'information clinique pertinente (données PMSI, biologies, historiques de soins/prescriptions...)
- Tout document relatif aux patients ayant exercé leur droit d'opposition à la réutilisation de leurs données de santé

Critères d'inclusion :

- Tout document produit ou intégré dans Sillage signé numériquement et ne répondant à aucun critère de non-inclusion.
- Document rattaché à tout séjour dont la date d'entrée se situait entre le 1^{er} janvier et le 31 janvier inclus, et le 1^{er} juillet et le 31 juillet inclus, sur l'année 2018 et 2023.
- Document rattaché à au moins une des unités fonctionnelles (UF) suivantes :
 - Ensemble des urgences adultes et enfants : 308*, 309*
 - Médecine interne et infectiologie : 221*, 222*, 223*, 230*
 - Pneumologie : 529*, 530*, 531*
 - Pédiatrie : 420* à 425*

Critère d'exclusion :

- Sur la base du corpus initial constitué selon les critères précédents, exclusion des documents ne mentionnant aucun terme recherché (lexique détaillé en 3.2.1).

Afin de garantir une sensibilité maximale, la stratégie adoptée privilégiait la non-inclusion des documents non pertinents plutôt que l'inclusion directe des documents pertinents, trop complexe et peu fiable (plusieurs dénominations pour un même type de documents, mauvais libellés).

3.2 Requête dans l'EDS et requête textuelle

On définissait par « **requête EDS** » l'étape consistant à extraire, à partir de l'EDS, un corpus initial de documents susceptibles de contenir le lexique d'intérêt. Cette première requête correspondait directement à l'application des critères d'inclusion précédemment décrits : elle s'appuyait exclusivement sur les caractéristiques médico-administratives associées au document, appelées « métadonnées », et non sur le contenu du document lui-même (identifiants, sexe et âge du patient, date d'entrée du séjour, code UF du document...).

La requête EDS avait pour finalité de générer une table contenant le document, stocké dans une unique chaîne de caractères, accompagné de ses métadonnées regroupées en 4 catégories thématiques :

- Identifiants : document, patient, séjour
- Caractéristiques « patient » : sexe, âge, dates de naissance et de décès
- Caractéristiques « séjour » : dates d'entrée et de sortie, durée de séjour
- Caractéristiques « document » : type et libellé du document, UF d'origine

On définissait par « **requête textuelle** » l'étape consistant à extraire, à partir du corpus initial, un corpus final de documents contenant le lexique d'intérêt. Cette seconde requête correspondait directement à l'application du critère d'exclusion précédemment décrit : elle s'appuyait exclusivement sur le contenu du document lui-même.

3.2.1 Terminologies recherchées

La requête textuelle était structurée autour de deux principes : l'utilisation d'expressions régulières pour l'identification du lexique, et l'organisation du lexique identifié en concepts.

Une « **expression régulière** » (ou « **regex** ») est une séquence de caractères définissant un motif textuel selon une syntaxe formelle standardisée. L'utilisation de tels motifs permet une recherche flexible intégrant l'ensemble des variantes lexicales via l'usage de métacaractères, par exemple :

- générique : . (« tout caractère »), \w (« tout caractère alphanumérique »), \s (« un espace »)
- quantificateurs : + (« au moins un »), * (« 0 ou plus »)
- classes : [a-z] (« toutes les lettres minuscules entre a et z »)
- ancrés : ^ (« commence par »), \$ (« finit par »)

Un « **concept** » désigne une entité abstraite pouvant être exprimée par une multitude de variantes lexicales. Chaque concept est traduit par un ensemble de regex capturant ses différentes manifestations textuelles : dénominations multiples, synonymes, abréviations, mais aussi fautes de frappe et variantes grammaticales.

Par exemple, le concept « érysipèle » peut apparaître dans les textes médicaux sous diverses formes :

- Avec orthographe correcte : « érysipèle »
- Avec fautes de frappe courantes : « érysipele », « erisipèle »
- Avec erreur phonétique courante : « érésipèle »
- Sous sa dénomination plus moderne : « dermo-hypodermite bactérienne non nécrosante », « DHBNN »

La formulation présentée en Figure 1 permet de capturer la majorité des variantes regroupées sous le même concept « érysipèle ».

```
(?i)[eé]r[yié]sip[eè]le?|dermo[\s-\s]?hypod[a-z]+|dhbnn
```

Figure 1. Exemple de recherche d'expressions régulières couvrant le concept « érysipèle ».

Recherche du mot « érysipèle » en incluant ses principales variantes orthographiques, ainsi que dermo (-) hypodermite (sans qualificatifs pour simplifier l'exemple).

L'approche par concept garantit l'exhaustivité de la requête textuelle permise par l'utilisation de regex tout en maintenant une structure sémantique.

Dans un premier temps, le lexique était organisé selon trois catégories :

- **Les AINS :**
 - « AINS » ou « anti-inflammatoires non stéroïdien » sans précision
 - Dénominations communes internationales (DCI) : ibuprofène, kétoprofène, naproxène, diclofénac, acéclofénac, célecoxib, méloxicam, piroxicam
 - Noms de spécialités commerciales : Nurofen®, Advil®, Antarène®, Spifen®, Profenid®, Biprofenid®, Voltarène®
- **Les typologies d'infections :**
 - abcès, empyème, cellulite, fasciite
 - angine, pharyngite, syndrome de Lemierre
 - otite, sinusite, mastoïdite, ethmoïdite, conjonctivite
 - méningite, méningo-encéphalite, épidurite
 - pneumonie, pneumopathie, pleuro-pneumopathie
 - érysipèle, dermohypodermité
 - varicelle
 - sepsis, choc
- **Les agents pathogènes :** noms vernaculaires ou scientifiques
 - streptocoque, *Streptococcus*
 - staphylocoque, *Staphylococcus*
 - pneumocoque, méningocoque
 - *Pseudomonas*

La finalité de la requête était que seuls soient retenus les documents mentionnant, dans le même document, une exposition à un AINS et un contexte infectieux en cours. Ainsi, les trois catégories initiales ont été simplifiées avec regroupement des typologies infectieuses et agents pathogènes, formant deux groupes opérationnels de concepts (Figure 2) :

- Un groupe « **AINS** », qui renferme 1 concept par molécule avec regroupement des DCI et spécialités commerciales (p ex., les termes « ibuprofène », « Nurofen » et « Advil » étaient regroupés sous un même concept « ibuprofène »).
- Un groupe « **PATHO** », qui renferme les sous-catégories correspondant aux typologies infectieuses et aux agents pathogènes, chacune renfermant 1 concept par entité ou groupe d'entités (p ex., les termes « otite », « sinusite », « mastoïdite » et « ethmoïdite » étaient regroupés sous un même concept « ORL »).



Figure 2. Groupe de concepts utilisés pour la requête textuelle.

La stratégie de recherche adoptée privilégie délibérément une maximisation de la sensibilité au détriment de la spécificité, en suivant le principe selon lequel l'omission d'un cas d'intérêt est plus préjudiciable que l'inclusion à tort d'un cas non pertinent. Cette approche se caractérise par deux choix méthodologiques assumés :

- Une recherche lexicale sans analyse contextuelle, qui ne distingue pas les affirmations des négations (« a pris un ibuprofène hier soir » vs « pas d'AINS »), ni les antécédents d'un épisode en cours. Cette simplification garantissait qu'aucune mention pertinente n'était écartée sur des critères linguistiques potentiellement mal définis.
- Une définition volontairement inclusive des entités pathologiques. Un terme comme « cellulite » était recherché sans restriction anatomique (dentaire, orbitaire, de membre...). Un terme comme « angine » était recherché sans restriction étiologique (bactérienne ou virale). De plus, l'objectif étant *in fine* d'identifier des cas d'infections ou de surinfections bactériennes, la recherche d'entités exclusivement d'origine virale au départ était justifiée (p ex., varicelle). Ce manque de précision visait à faciliter la probabilité de capturer des présentations cliniques atypiques ou mal définies.

On postulait qu'une telle requête, bien que générant une proportion importante de faux positifs, présenterait toujours une meilleure probabilité de contenir des cas d'intérêt comparativement à un échantillonnage aléatoire dans l'EDS. Les regex étaient formulées et validées à dires d'expert, individuellement pour chaque terme recherché puis regroupées par concepts.

3.2.2 Extraction du texte

Une extraction intermédiaire était réalisée à partir du corpus initial selon une logique d'union booléenne, ne retenant que les documents satisfaisant la condition suivante :

Présence d'au moins un concept du groupe « AINS » OU du groupe « PATHO »

Les documents ne respectant pas cette condition étaient exclus. Enfin, une extraction finale était réalisée à partir de ce corpus intermédiaire selon une logique d'intersection booléenne, ne retenant que les documents satisfaisant la condition suivante :

Présence d'au moins un concept du groupe « AINS » ET du groupe « PATHO »

Les documents ne respectant pas cette condition étaient exclus.

Cette extraction était réalisée en deux temps pour description dans le diagramme de flux, dont les étapes successives sont présentées en Figure 3. Une étape finale de dédoublonnage (au niveau du patient) était finalement appliquée par sélection, pour chaque patient, du premier document signé numériquement. Ainsi le terme « population » était employé pour qualifier indistinctement le groupe de patients, de documents ou même de séjours auxquels ces derniers étaient rattachés.

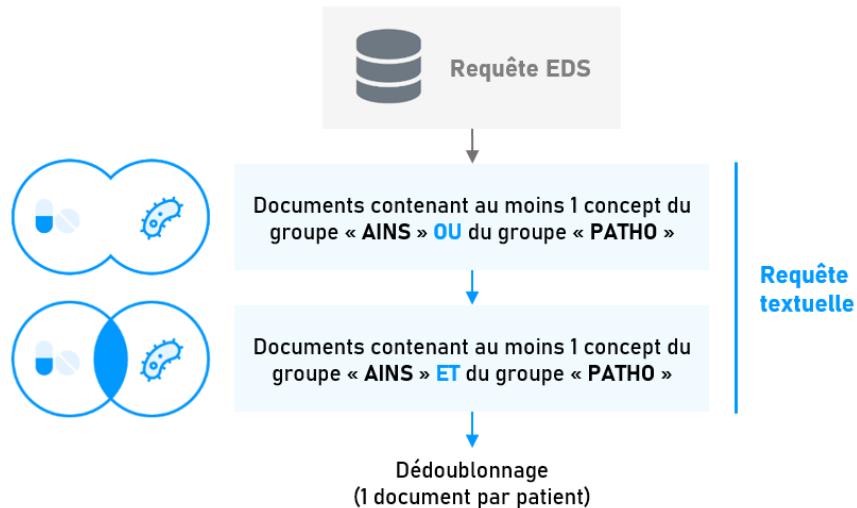


Figure 3. Requête EDS et requête textuelle pour constitution d'un corpus de documents à annoter.

Ce corpus final de documents constituait la population sur laquelle portaient les différentes phases d'annotations.

4 Annotation experte : classification de référence

On définissait par « **annotation experte** » le processus d'analyse des documents réalisé par un ou plusieurs lecteurs humains. Elle s'opposait à l'annotation dite « automatisée », qui correspond à ce même processus réalisé par un lecteur non-humain. L'objectif de l'annotation experte était de classer les cas selon leur éligibilité à une analyse d'imputabilité (non réalisée dans le cadre de cette étude).

4.1 Classification des cas

Le corpus de documents à annoter était constitué de deux ensembles distincts :

- L'intégralité des documents produits en janvier et juillet 2018.
 - Un échantillon aléatoire tiré des documents produits en janvier et juillet 2023, dont la taille était calculée sur la base de l'hypothèse suivante : l'estimation d'une *accuracy* attendue de 70 %, avec une précision de $\pm 7,5$ % et un risque de première espèce égal à 5%. Ce calcul nous a permis d'estimer cette taille d'échantillon à 143.

L'annotation experte visait à identifier les cas d'infections bactériennes survenues ou aggravées dans un contexte d'exposition confirmée ou suspectée à un AINS préalablement à l'admission. L'unité d'analyse étant le couple patient-document, le terme de « cas » pouvait être employé pour désigner indistinctement le document annoté ou le patient auquel il était

rattaché. Chaque cas était classé de façon binaire en « éligible » ou « inéligible ». Un cas classé éligible était un cas jugé susceptible de faire l'objet d'une déclaration en pharmacovigilance et justifiant ainsi une analyse d'imputabilité. Cette analyse n'était pas réalisée dans le cadre de cette étude.

Pour être classé éligible, un cas devait obligatoirement satisfaire les 3 critères suivants :

1. L'infection ou ses symptômes précurseurs constituait le motif principal d'admission à l'hôpital.
2. La nature bactérienne de l'infection était confirmée par la présence d'au moins 1 critère parmi les suivants :
 - Prescription d'une antibiothérapie, et/ou
 - Réalisation d'un drainage chirurgical, et/ou
 - Identification de l'agent pathogène
3. La dernière prise d'AINS documentée devait être récente.

L'annotation était réalisée en double lecture indépendante par deux pharmacologues en pharmacovigilance. La mesure du coefficient kappa de Cohen, permettant de quantifier la concordance inter-expert ajustée sur le hasard, était réalisée. Les cas de discordance faisaient l'objet d'une révision conjointe jusqu'à obtention d'un consensus.

La finalité de cette classification était de mesurer la sensibilité de la requête et constituer une référence (*gold standard*) pour la formulation des instructions et évaluation des capacités discriminantes du LLM pour l'étape d'annotation automatisée.

4.2 Processus d'annotation

L'annotation experte était réalisée via une interface développée avec l'outil de création d'applications web *Streamlit* (version 1.44.0, environnement Python 3.11.11), permettant une visualisation optimisée et une saisie structurée des données. Cette interface offrait un environnement standardisé garantissant l'uniformité et l'indépendance du processus d'annotation entre les différents experts. Elle était structurée en deux espaces fonctionnels distincts : un espace de visualisation du document et un espace de saisie des annotations.

D'une part, le document était mis à disposition dans sa forme intégrale et native (tel qu'il se présente dans le SI utilisé en routine par le clinicien), avec surbrillance du lexique d'intérêt. Cette fonctionnalité, rendue possible par la requête textuelle, permettait une localisation rapide des passages pertinents tout en préservant la lecture du contexte global. Une contextualisation supplémentaire était proposée en présentant en regard du texte :

- Ses principales métadonnées, directement extraites de l'EDS : identifiants, caractéristiques patient (âge, sexe), données du séjour (dates d'entrée/sortie), caractéristiques document (libellé, UF d'origine)
- Le détail du lexique d'intérêt, extrait par la requête textuelle : concepts identifiés dans le texte + extraction d'un verbatim des termes identifiés avec respect du nombre d'occurrence et de l'ordre d'apparition dans le texte.

D'autre part, un espace de saisie avec sélecteur à boutons radio permettait de classer le cas en « éligible » ou « inéligible » selon les critères précédemment définis.

Pour la navigation dans l'interface, l'index du document permettait de visualiser la progression dans le corpus et l'accès à un document précis par saisie directe, facilitant les révisions ou la reprise après interruption. Des boutons « précédent/suivant » permettaient une navigation bidirectionnelle dans le corpus avec enregistrement de la saisie. Les annotations étaient enregistrées dans une table stockée sur un serveur distant, inaccessible à l'utilisateur. Une capture écran de l'interface d'annotation est présentée en Figure 4.

L'interface d'annotation experte (AINS_2018) est un formulaire en ligne. En haut à gauche, il indique "AINS_2018", "Document 51 sur 677". En haut à droite, des informations patient sont affichées : ID DOC, ID PAT, ID SEJ, AGE : 34.7 – SEXE : M, DATE ENTRÉE : 2018-07-29, DATE SORTIE : 2018-07-31, UF : 3095 - HCD URGENCE CHIR ET SOINS PERI OP, TITRE : CR examen RESURGENCE. Le formulaire contient des sections pour "Retenu lecteur 1" (Commentaire lecteur 1), "Retenu lecteur 2" (Commentaire lecteur 2), "Retenu final" (Raison du désaccord / justification). Des boutons "ENREGISTRER" et "CONCEPTS" sont disponibles. La section "CONCEPTS" montre une liste de termes : "abcès ; cellulite ; ibuprofène". La section "TEXTE" montre une liste de termes : "abcès ; abcès ; IBUPROFENE ; abcès ; abcès ; ibuprofène ; cellulite ; abcès ; abcès ; Abcès ; abcès". À droite, une grande zone de texte détaille l'origine du patient (SERVICE URGENCE AUTRES URG), le motif d'admission (Anamnèse), le motif d'entrée (abcès retro pharyngé droit), le motif de venue (Or: Douleur pharyngée), les circonstances (Autre), la famille (Présente), l'observation médicale (structuree), le motif de recours (TF Seclin abcès rétropharyngien droit), le traitement habituel (0), l'anamnèse (douleur amygdale droite depuis 48 heures), l'écriture (a pris un IBUPROFENE hier soir), la consultation (Consulte à Seclin pour difficulté avaler + douleur intense), l'envoi à CHRU (Envoyer CHRU pour PEC ORL d'un abcès rétropharyngé droit), les examens cliniques (Examen clinique apyrétique), l'état (HD stable), les douleurs (Oedème sous-mandibulaire droit, douloureux à la palpation), les trismes (Trismus à 2 doigts), la déviation de la lèvre (Déviation de la lèvre, érythème et bombement de la paroi pharyngée droite), les douleurs (Douleurs importante à la déglutition devenue difficile même pour la salive), les dents (Dents non examinables au vu du trismus), la synthèse (Synthèse des examens complémentaires et avis spé Bio : Avis ORL : n'a pas été prévenu par Seclin du transfert passe à 7h, essaiera de ponctionner, pas de TDM d'emblée Conclusion et Orientation Phlegmon rétropharyngé droit).

Figure 4. Capture écran de l'interface mise à disposition pour l'annotation experte (classification des cas). Ici formatée pour résolution des désaccords inter-experts (capture datée de 06/2025).

5 Annotation automatisée : stratégie de détection

On définissait par « **annotation automatisée** » le processus d'analyse des documents réalisé par un lecteur non-humain. Elle s'opposait à l'annotation experte décrite précédemment. L'annotation automatisée était réalisée par un LLM et poursuivait un double objectif :

1. Classification des cas avec développement et optimisation de la stratégie de détection (*prompt engineering*)
2. Évaluation de la capacité discriminante de cette stratégie

Le corpus de documents soumis à l'annotation automatisée était identique à celui soumis à l'annotation experte. Dans le but de tester la stratégie de détection, ce corpus était stratifié en deux ensembles distincts :

- L'intégralité des documents produits en janvier et juillet 2018 constituait un ensemble « **d'apprentissage** », utilisé pour la formulation et l'optimisation itérative des instructions données au LLM.
- L'échantillon aléatoire tiré des documents produits en janvier et juillet 2023 constituait un ensemble de « **test** », réservé à l'évaluation des capacités discriminantes du LLM sur de nouvelles données.

5.1 Modèle de langue et paramètres

L'annotation automatisée était réalisée par le modèle Llama version 3.3 à 70 milliards de paramètres (Llama3.3:70B), LLM à architecture ouverte développé par Meta et déployé en décembre 2024⁶⁰. Ce modèle généraliste, unimodal, était choisi pour la possibilité d'une implémentation locale, garantissant la confidentialité des données médicales, l'absence de restrictions d'usage ainsi qu'un contrôle total des paramètres d'inférence. De plus, des performances supérieures aux architectures comparables étaient observées de façon empirique sur le traitement contextuel de textes médicaux en langue française. L'implémentation locale du LLM était réalisée via l'interface *open source* Ollama (version 0.11.4), facilitant le paramétrage et l'exécution locale des LLM à architecture ouverte et l'optimisation des ressources. Elle garantissait une exécution exclusivement locale, sans transmission de données vers des serveurs externes. Deux paramètres d'inférence du LLM étaient spécifiquement ajustés dans le cadre de cette étude :

- La « **température** », paramètre contrôlant la variabilité des réponses en modifiant la distribution de probabilité de tous les tokens possibles, avec des valeurs comprises entre 0 et 1. Une température de 0 rend le LLM déterministe en sélectionnant systématiquement le token le plus probable. Une température de 0,8 (valeur par défaut) introduit une variabilité contrôlée, permettant d'évaluer la robustesse des réponses face aux variations stochastiques.
- Le « **top-p** », paramètre de filtrage limitant l'espace de sélection aux tokens les plus probables dont la somme cumulée atteint le seuil p (entre 0 et 1). Un top-p de 0 limite la sélection au token le plus probable, tandis qu'un top-p de 0,9 (valeur par défaut) permet au LLM de considérer un ensemble plus large de tokens possibles.

L'ajustement conjoint de ces deux paramètres permet de contrôler le degré de créativité (ou de prévisibilité) du LLM. Afin de maximiser les chances d'obtenir une réponse parfaitement structurée et déterministe, température et top-p étaient fixés à 0 (configuration dite « déterministe »). Pour évaluer l'impact d'un tel paramétrage, l'annotation était reproduite sur le même corpus en laissant à ces paramètres leur valeur par défaut (configuration dite « stochastique »).

Quelle que soit sa configuration, le LLM traitait chaque document individuellement, sans mémorisation des textes précédents, sans accès aux données de l'annotation experte ni à tout autre contexte externe. L'annotation automatisée était réalisée en suivant des instructions précisément formulées et nécessitant une phase d'optimisation.

5.2 Formulation initiale des instructions

L'étape de *prompt engineering* était réalisée sur l'ensemble d'apprentissage. Il convient de préciser que cette phase ne relevait pas d'un apprentissage du modèle au sens traditionnel du terme employé en *machine learning*, mais d'une optimisation empirique des instructions visant à maximiser les capacités discriminantes du LLM.

Ce processus suivait une démarche itérative structurée. Dans un premier temps, une formulation initiale du prompt était testée sur un échantillon de documents pour identifier les erreurs récurrentes de classification : l'analyse qualitative de ces erreurs permettait d'identifier les ambiguïtés et les critères mal interprétés par le LLM. Les instructions étaient alors

reformulées, précisées ou réorganisées, puis testées à nouveau jusqu'à stabilisation des capacités discriminantes, mesurées en termes de sensibilité et de valeur prédictive positive. Ce cycle était répété à environ 5 reprises sur l'ensemble d'apprentissage. L'ensemble de test n'était pas utilisé pour cette phase d'optimisation.

Le prompt finalement constitué était structuré en cinq sections explicitement nommées :

A	IDENTITÉ	Attribution d'un rôle au LLM
B	OBJECTIF	Définition de la tâche de classification
C	CRITÈRES D'ÉLIGIBILITÉ	Énonciation des critères de classification des cas en éligible ou inéligible (identiques aux critères définis pour l'annotation experte)
D	RAPPELS	Principaux acronymes, liste non exhaustive d'entités à inclure ou exclure
E	RÉPONSE ATTENDUE	Spécification du format de réponse + instructions complémentaires visant à réduire les erreurs récurrentes et facilement évitables

Le format de réponse utilisé était le JSON (*JavaScript Object Notation*), structure de données standardisée associant chaque réponse (critères d'éligibilité et classification finale) à sa valeur booléenne (condition logique true/false). Cette standardisation garantissait une extraction déterministe et uniforme des résultats, facilitant leur agrégation sur l'ensemble du corpus et leur manipulation en vue des analyses ultérieures.

Le détail simplifié de la réponse demandée pour chaque texte était le suivant :

presence_infection	« Le texte fait-il mention d'une admission à l'hôpital pour une infection en cours ou des symptômes précurseurs d'infection ? »
infection_bacterienne	« La nature bactérienne de l'infection est-elle confirmée ? »
exposition_ains	« Le texte fait-il mention d'une exposition à un AINS préalablement à la venue à l'hôpital ? »
cas_eligible	« Sur la base de tes réponses, détermine si le cas répond aux critères d'éligibilité »

La structuration de la réponse, exigeant une évaluation explicite de chaque critère avant la classification du cas, permettait une estimation de la cohérence interne du LLM. Deux types de classification pouvaient être distingués :

- Une classification dite « **déclarée** », correspondant à la classification des cas effectivement produite par le LLM (champ « **cas_eligible** »).
- Une classification dite « **théorique** », résultant de l'application stricte de la règle logique formulée dans le prompt : un cas ne pouvait être classé « éligible » que si et seulement si les trois critères étaient satisfaits.

La mise en regard de la classification déclarée avec la classification théorique correspondante permettait d'identifier les incohérences logiques du LLM, situations où la classification du cas contredisait l'évaluation individuelle des critères.

5.3 Plan d'optimisation finale des instructions

Plusieurs déclinaisons dans le format du prompt ainsi constitué étaient produites et testées systématiquement selon un plan prédéfini, permettant d'évaluer l'impact de variations contrôlées sur les capacités discriminantes du LLM. Dans un premier temps, le contenu du prompt était formaté de deux manières différentes, donnant naissance à deux modèles de prompts dont la distinction fondamentale était la suivante :

Modèle « 0 »	Les différentes sections étaient rédigées en langage naturel, afin de privilégier la lisibilité (sauf exception)
Modèle « 1 »	Les différentes sections étaient rédigées sous forme de dictionnaire JSON, afin de privilégier la standardisation et la concision (sauf exception).

Il convient de préciser que cette dichotomie portait exclusivement sur la forme du prompt, et non sur la nature des instructions qui le composent. Il y avait deux exceptions à cette règle : l'une concernait l'attribution du rôle (section A), qui était simplement supprimée dans le modèle 1 ; l'autre concernait le format de réponse (section D), qui possédait dans les deux cas une structure de base en dictionnaire JSON, avec des spécificités propres à chaque modèle. Les spécificités de chaque modèle de prompt sont présentées en Annexes 1 et 2.

Afin d'estimer l'impact du formatage du prompt sur les capacités discriminantes du LLM, un ensemble de prompts dérivés était généré selon un plan factoriel croisant individuellement les cinq sections (A, B, C, D, E) dans leurs deux formats possibles (modèle 0 ou 1), produisant ainsi $2^5 = 32$ prompts distincts. L'annotation automatisée de l'ensemble d'apprentissage était produite de façon itérative avec les 32 prompts, dans chacun des deux paramétrages du LLM décrits précédemment (déterministe et stochastique), totalisant un nombre final de 64 annotations. Ainsi, à chaque annotation automatisée correspondait une configuration unique associant un prompt et un paramétrage du LLM.

6 Description des cas éligibles : analyse exploratoire

À titre exploratoire, les cas classés éligibles par la référence faisaient l'objet d'une description visant à documenter les éléments nécessaires à une analyse d'imputabilité. Cette étape descriptive était réalisée par un unique lecteur, sur le même mode que pour l'annotation experte (interface présentée en 4.2). Un espace de saisie était spécifiquement configuré : formulaire détaillé combinant boutons radio pour les variables binaires (p ex. contexte de prise, prescription d'une antibiothérapie) et champs de saisie textuelle pour les données nécessitant une extraction littérale (p ex. chronologie, AINS consommé, agent pathogène identifié).

La grille de caractéristiques à documenter était constituée des éléments suivants :

- Données relatives à l'exposition à un AINS :
 - Molécule impliquée, en DCI
 - Temps écoulé entre la dernière prise d'AINS documentée et l'admission
 - Contexte de prise : automédication ou prescription médicale
 - Antibiothérapie associée avant l'admission

- Données relatives à l'épisode infectieux en cours :
 - Temps écoulé entre le début déclaré des symptômes et l'admission
 - Infection bactérienne diagnostiquée
 - Prescription d'une antibiothérapie
 - Réalisation d'un drainage chirurgical
 - Identification d'un agent pathogène
 - Mention explicite de l'arrêt ou d'une contre-indication à l'AINS par le clinicien
 - Évolution (statut de sortie)

L'absence d'information était classée en donnée manquante et non incluse dans le calcul des proportions.

Parallèlement, sur l'ensemble de la période étudiée, une recherche de cas déclarés au CRPV du CHU de Lille en lien avec les pathologies identifiés était réalisée.

7 Analyses statistiques

Les variables catégorielles et binaires étaient présentées en termes d'effectifs avec pourcentages. En la présence de distributions asymétriques, les variables quantitatives étaient présentées en termes de médianes avec intervalles interquartiles.

L'évaluation des capacités discriminantes du LLM était réalisée distinctement sur les ensembles d'apprentissage et de test. Chaque annotation automatisée était confrontée à l'annotation experte, utilisée comme référence. Plusieurs métriques de classification étaient calculées :

- La **sensibilité**, ou **rappel (Se)**, proportion de cas correctement classés éligibles par l'annotation automatisée parmi l'ensemble des cas réellement éligibles.
- La **valeur prédictive positive**, ou **précision (VPP)**, proportion de cas réellement éligibles parmi l'ensemble des cas classés éligibles par l'annotation automatisée.
- La **F-mesure**, moyenne harmonique entre sensibilité et VPP (F1).
- La **spécificité (Sp)**, proportion de cas correctement classés inéligibles par l'annotation automatisée parmi l'ensemble des cas réellement inéligibles.

Parmi les 64 annotations automatisées réalisées sur l'ensemble d'apprentissage, la configuration associée (prompt + paramétrage du LLM) qui maximisait le score F1, selon la classification déclarée, était sélectionnée pour application sur l'ensemble de test.

Les capacités discriminantes du LLM sur l'ensemble d'apprentissage étaient présentées à la fois vis-à-vis de la classification déclarée et théorique. Concernant l'ensemble de test, seuls les résultats vis-à-vis de la classification déclarée étaient présentés. Les métriques précédemment décrites étaient calculées pour chaque ensemble.

Toutes les analyses statistiques étaient réalisées avec R version 4.5.1.

Résultats

1 Population

1.1 Diagramme de flux

Sur la période d'étude, 99 573 documents répondaient aux critères d'inclusion et de non-inclusion définis par la requête EDS. Parmi ces documents, la requête textuelle identifiait 22 208 (22,3 %) contenant au moins un terme recherché tout concept confondu, parmi lesquels 1 318 (1,3 %) contenaient une co-occurrence d'au moins un concept du groupe « AINS » et du groupe « PATHO ». Après dédoublonnage, 1 198 (1,2 %) documents étaient finalement retenus pour annotation. Parmi ces documents, 626 (52,3 %) étaient produits sur la période 2018, 572 (47,7 %) étaient produits sur la période 2023. Un échantillon aléatoire de 143 documents était tiré à partir des documents produits sur la période 2023.

Le diagramme de flux est représenté en Figure 5.

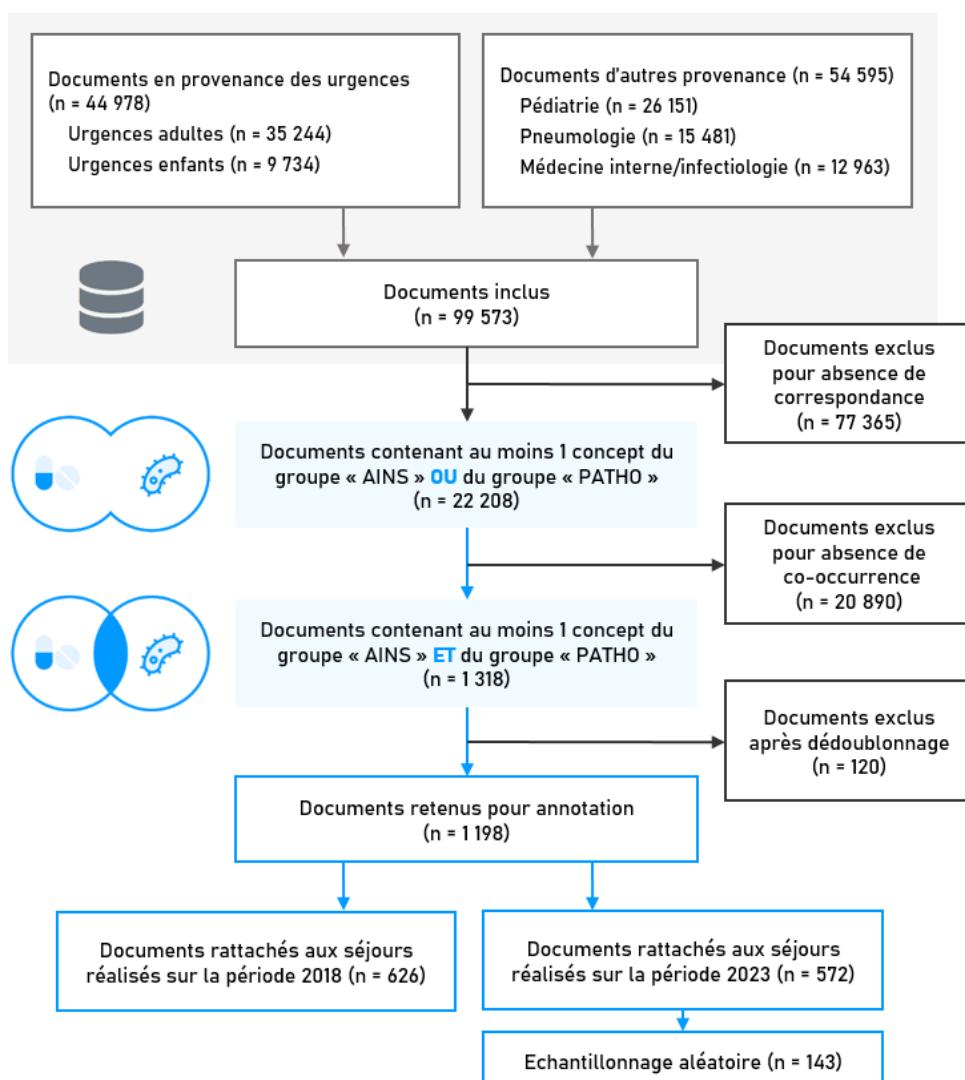


Figure 5. Diagramme de flux.

1.2 Description de la population retenue pour annotation

L'ensemble des 626 documents produits sur la période 2018, ainsi que les 143 documents échantillonnés sur la période 2023, étaient conservés pour annotation. Les effectifs de la population finalement retenue pour annotation sont présentés en Table 1.

Table 1. Descriptif de la population retenue pour annotation, globalement et selon l'année.

Variable	Total (N=769)	Année	
		2018 (n=626, 81,4%)	2023 (n=143, 18,6%)
Date d'entrée du séjour, n (%)			
1er au 31 janvier inclus	433 (56,3)	352 (56,2)	81 (56,6)
1er au 31 juillet inclus	336 (43,7)	274 (43,8)	62 (43,4)
Nature du document analysé, n (%)			
Compte rendu urgences	570 (74,1)	467 (74,6)	103 (72,0)
Lettre de sortie	99 (12,9)	75 (12,0)	24 (16,8)
Compte rendu de consultation	82 (10,7)	70 (11,2)	12 (8,4)
Divers	18 (2,3)	14 (2,2)	4 (2,8)

Nombre de documents. Chaque document était rattaché à un patient distinct.

2 Annotation experte : classification de référence

Au total, sur les 769 documents analysés, 122 (15,9 %) étaient classés éligibles après consensus, comme présenté en Figure 6.

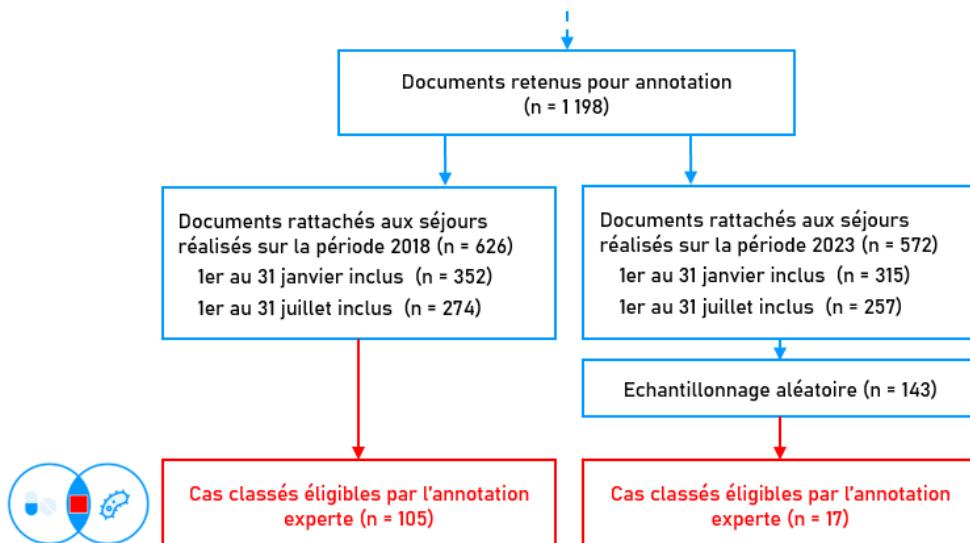


Figure 6. Nombre de cas classés éligibles par l'annotation experte après consensus, à partir de la population retenue pour annotation, sur l'ensemble de période d'étude.

Sur 626 documents annotés sur la période 2018, 105 (16,7 %) étaient classés éligibles après consensus entre les deux experts. Le coefficient kappa pour la concordance inter-expert avant consensus était de 75,0 %. Le tableau de contingence croisant l'annotation de chaque expert avant consensus est présenté en Table 2.

Table 2. Table de contingence croisant les annotations de chaque expert avant consensus, année 2018.

		Expert 1		Total
		Cas éligible	Cas inéligible	
Expert 2	Cas éligible	74 (11,8%)	31 (5,0%)	105 (16,8%)
	Cas inéligible	9 (1,4%)	512 (81,8%)	521 (83,2%)
	Total	83 (13,3%)	543 (86,7%)	626 (100,0%)

Chaque annotation était réalisée indépendamment l'une de l'autre. Le coefficient kappa pour la concordance inter-expert avant consensus était de 75,0%.

Sur 143 documents annotés sur la période 2023, 17 étaient classés éligibles après consensus entre les deux experts. Le coefficient kappa pour la concordance inter-expert avant consensus était de 85,9 %. Le tableau de contingence croisant l'annotation de chaque expert avant consensus est présenté en Table 3.

Table 3. Table de contingence croisant les annotations de chaque expert avant consensus, année 2023.

		Expert 1		Total
		Cas éligible	Cas inéligible	
Expert 2	Cas éligible	14 (9,8%)	2 (1,4%)	16 (11,2%)
	Cas inéligible	2 (1,4%)	125 (87,4%)	127 (88,8%)
	Total	16 (11,2%)	127 (88,8%)	143 (100,0%)

Chaque annotation était réalisée indépendamment l'une de l'autre. Le coefficient kappa pour la concordance inter-expert avant consensus était de 85,9%.

3 Annotation automatisée : stratégie de détection

3.1 Capacité discriminante sur l'ensemble d'apprentissage

La configuration sélectionnée, parmi les 64 exécutées sur l'ensemble d'apprentissage, associait un paramétrage déterministe du LLM (température et top-p fixés à 0) avec un prompt correspondant à la nomenclature « A0_B1_C0_D0_E0 ». Le contenu détaillé du prompt est présenté en Annexe 3.

3.1.1 Selon la classification déclarée

Parmi les 626 cas annotés par le LLM sur la période 2018, 130 étaient classés éligibles selon la classification déclarée. Parmi les 105 cas classés éligibles par l'annotation experte, 92 étaient classés éligibles selon la classification déclarée du LLM. Le croisement de l'annotation experte avec la classification déclarée du LLM rapportait une sensibilité de 87,6 %, une spécificité de 92,7 %, et une VPP de 70,8 %. La F1 était de 78,3 %.

Le croisement des résultats de l'annotation automatisée (classification déclarée) avec ceux de l'annotation experte est présenté en Table 4.

Table 4. Table de contingence croisant l'annotation experte avec l'annotation automatisée (classification déclarée), année 2018.

		Annotation experte		Total
		Cas éligible	Cas inéligible	
Annotation auto (CD)				
Cas éligible	92 (14,7%)	38 (6,1%)	130 (20,8%)	
Cas inéligible	13 (2,1%)	483 (77,2%)	496 (79,2%)	
Total	105 (16,8%)	521 (83,2%)	626 (100,0%)	

L'annotation automatisée correspond à la classification des cas réalisée par le LLM. La classification déclarée est la réponse donnée par le LLM, après appréciation globale du cas sur la base de ses réponses aux critères de sélection individuels. L'annotation experte est utilisée comme référence. CD : classification déclarée.

3.1.2 Selon la classification théorique

Sur les 626 cas annotés par le LLM sur la période 2018, 137 étaient classés éligibles selon la classification théorique. Parmi les 105 cas classés éligibles par l'annotation experte, 92 étaient classés éligibles selon la classification théorique du LLM. Le croisement de l'annotation experte avec la classification théorique du LLM rapportait une sensibilité de 87,6 %, une spécificité de 91,4 %, et une VPP de 67,2 %. La F1 était de 76,0 %.

Le croisement des résultats de l'annotation automatisée (classification théorique) avec ceux de l'annotation experte est présenté en Table 5.

Table 5. Table de contingence croisant l'annotation experte avec l'annotation automatisée (classification théorique), année 2018.

		Annotation experte		Total
		Cas éligible	Cas inéligible	
Annotation auto (CT)				
Cas éligible	92 (14,7%)	45 (7,2%)	137 (21,9%)	
Cas inéligible	13 (2,1%)	476 (76,0%)	489 (78,1%)	
Total	105 (16,8%)	521 (83,2%)	626 (100,0%)	

L'annotation automatisée correspond à la classification des cas réalisée par le LLM. La classification théorique est la réponse attendue du LLM, selon l'application stricte de la condition logique (le cas doit être classé éligible si et seulement si l'ensemble des critères de sélection sont satisfaits). L'annotation experte est utilisée comme référence. CT : classification théorique.

Les résultats des 64 annotations réalisées sur l'ensemble d'apprentissage sont présentés en Annexe 4.

3.2 Capacité discriminante sur l'ensemble de test

Sur l'échantillon de 143 cas annotés par le LLM sur la période 2023, 25 étaient classés éligibles selon la classification déclarée. Parmi les 17 cas classés éligibles par l'annotation experte, 15 étaient classés éligibles selon la classification déclarée du LLM. Le croisement de l'annotation experte avec la classification déclarée du LLM rapportait une sensibilité de 88,2 %, une spécificité de 92,1 %, et une VPP de 60,0 %. La F1 était de 71,4 %.

Le croisement des résultats de l'annotation automatisée (réponses observée) avec ceux de l'annotation experte est présenté en Table 6.

Table 6. Table de contingence croisant l'annotation experte avec l'annotation automatisée (classification déclarée), année 2023.

	Annotation experte		Total
	Cas éligible	Cas inéligible	
Annotation auto (CD)			
Cas éligible	15 (10,5%)	10 (7,0%)	25 (17,5%)
Cas inéligible	2 (1,4%)	116 (81,1%)	118 (82,5%)
Total	17 (11,9%)	126 (88,1%)	143 (100,0%)

L'annotation automatisée correspond à la classification des cas réalisée par le LLM. La classification déclarée est la réponse donnée par le LLM, après appréciation globale du cas sur la base de ses réponses aux critères de sélection individuels. L'annotation experte est utilisée comme référence. CD : classification déclarée.

4 Description des cas éligibles : analyse exploratoire

Sur les 122 cas classés éligibles par l'annotation experte, 68 (55,7 %) étaient de sexe masculin. Chez les adultes, l'âge médian (IQR) était de 36,7 ans (25,7—49,6). Chez les enfants, l'âge médian (IQR) était de 7,6 ans (3,5—13,2). Les passages aux urgences concernaient 109 (89,3 %) patients, contre 13 (10,7 %) provenant d'autres secteurs.

Le délai entre le début déclaré des symptômes et l'admission était documentée pour 96 (78,7 %) patients, dont une majorité (n=36, 37,5 %) déclarant celui-ci inférieur à 3 jours. Les complications infectieuses les plus fréquemment représentées étaient des cas d'infections bucco-dentaires (n=52, 42,6%), principalement des cellulites dentaires (n=26, 21,3 %) et des abcès dentaires (n=17, 13,9 %) ; des infections ORL (n=31, 25,4%), principalement des angines bactériennes (n=13, 10,7 %) et des otites externes ou moyennes aiguës (n=14, 11,5 %) ; des infections cutanées (n=22, 18,0 %), principalement des dermo-hypodermites (n=9, 7,4 %) et des abcès (n=8, 6,6 %).

Concernant la prise d'AINS, 68 (55,7 %) patients ont déclaré avoir consommé de l'ibuprofène préalablement à l'admission, contre 38 (31,1 %) mentionnant un « AINS » sans précision et 16 (13,1 %) une autre molécule (kétoprofène, n=7 ; diclofénac, n=6 ; acide tiaprofénique, n=1 ; célecoxib, n=1 ; piroxicam, n=1). Le caractère ponctuel ou répété de la prise était insuffisamment documenté. Le temps écoulé entre la dernière prise déclarée et l'admission était documentée pour 86 (70,5 %) patients, dont une majorité (n=43, 50,0 %) déclarant celle-ci inférieure à 1 jour. La prise faisait l'objet d'une prescription médicale pour 57 (54,3 %) patients. Pour 44 (36,1 %) patients avaient une antibiothérapie associée avant l'admission. Le motif de prise le plus fréquent était la douleur pour 72 (59,0 %) patients (douleurs dentaires, migraines ou autres douleurs). Pour 40 (32,8%) patients, l'AINS était pris en contexte de syndrome infectieux (en particulier en présence d'angine, syndrome de pseudo-grippal ou d'otite). Enfin, 5 (4,1 %) cas concernaient des traitements chroniques et 5 (4,1 %) cas des motifs non renseignés.

Concernant la prise en charge, la prescription d'une antibiothérapie ou la poursuite d'une antibiothérapie préexistante était indiquée pour 111 patients (91,7 %). L'arrêt ou la contre-indication de l'AINS était explicitement mentionné par le médecin pour 33 (27,0 %) patients. Le retour à domicile était indiqué pour 91 (74,6 %) patients, tandis que 31 (25,4 %) étaient transférés en hospitalisation. La durée médiane de séjour (IQR) était de 13,8 heures (9,2—30,3). Le descriptif des cas classés éligibles par l'annotation experte est présenté en Table 7.

Table 7. Descriptif des cas classés éligibles par l'annotation experte après consensus, globalement et selon l'année.

Variable	Total (N=122)	Année	
		2018 (n=105, 86,1%)	2023 (n=17, 13,9%)
Sexe masculin, n (%)	68 (55,7)	58 (55,2)	10 (58,8)
Âge (adultes), années, médiane (IQR) ^a (N=107)	36,7 (25,7—49,6)	36,1 (25,1—49,4)	42,1 (27,5—53,5)
Âge (enfants), années, médiane (IQR) (N=15)	7,6 (3,5—13,2)	5,2 (3,5—13,2)	10,0 (8,6—11,4)
Âge, années, n (%)			
< 15,3	15 (12,3)	13 (12,4)	2 (11,8)
15,3 à 39,9	63 (51,6)	58 (55,2)	5 (29,4)
40 à 64,9	37 (30,3)	28 (26,7)	9 (52,9)
>= 65	7 (5,7)	6 (5,7)	1 (5,9)
Date d'entrée du séjour, n (%)			
1er au 31 janvier inclus	64 (52,5)	56 (53,3)	8 (47,1)
1er au 31 juillet inclus	58 (47,5)	49 (46,7)	9 (52,9)
Durée du séjour, heures, médiane (IQR)	13,8 (9,2—30,3)	13,9 (9,5—53,3)	13,3 (9,2—20,7)
Lieu d'hospitalisation, n (%) ^b			
Urgences	109 (89,3)	93 (88,6)	16 (94,1)
Autre	13 (10,7)	12 (11,4)	1 (5,9)
Exposition à un AINS			
Molécule impliquée, n (%)			
Ibuprofène	68 (55,7)	58 (55,2)	10 (58,8)
Kétoprofène	7 (5,7)	6 (5,7)	1 (5,9)
Diclofénac	6 (4,9)	6 (5,7)	0
Célécoxib	1 (0,8)	1 (1,0)	0
Piroxicam	1 (0,8)	0	1 (5,9)
Acide tiaprofénique	1 (0,8)	1 (1,0)	0
AINS sans précision	38 (31,1)	33 (31,4)	5 (29,4)
Délai prise-admission, jours, n (%) ^c (N=86)			
< 1	43 (50,0)	36 (47,4)	7 (70,0)
1 à 2	32 (37,2)	29 (38,2)	3 (30,0)
>= 2	11 (12,8)	11 (14,5)	0
Contexte de prise, n (%) ^d (N=105)			
Prescription médicale	57 (54,3)	51 (58,0)	6 (35,3)
Automédication	48 (45,7)	37 (42,0)	11 (64,7)
Délai symptômes-admission, jours, n (%) ^e (N=96)			
< 3	36 (37,5)	31 (39,2)	5 (29,4)
3 à 7	30 (31,2)	22 (27,8)	8 (47,1)
>= 7	30 (31,2)	26 (32,9)	4 (23,5)
Infection bactérienne diagnostiquée, n (%)			
Infection bucco-dentaire	52 (42,6)	42 (34,4)	10 (8,2)
Cellulite dentaire	26 (21,3)	20 (19,0)	6 (35,3)
Abcès dentaire	17 (13,9)	15 (14,3)	2 (11,8)
Abcès amygdalien	8 (6,6)	6 (5,7)	2 (11,8)

Amygdalite aiguë	1 (0,8)	1 (1,0)	0
Infection ORL	31 (25,4)	27 (22,1)	4 (3,3)
Angine bactérienne	13 (10,7)	10 (9,5)	3 (17,6)
Otite externe aiguë	9 (7,4)	9 (8,6)	0
Otite moyenne aiguë	5 (4,1)	4 (3,8)	1 (5,9)
Sinusite aiguë	3 (2,5)	3 (2,9)	0
Cellulite orbitaire	1 (0,8)	1 (1,0)	0
Infection cutanée	22 (18,0)	19 (15,5)	3 (2,5)
Dermo-hypodermite	9 (7,4)	9 (8,6)	0
Abcès cutané	8 (6,6)	5 (4,8)	3 (17,6)
Abcès péri-prothétique de hanche	2 (1,6)	2 (1,9)	0
Kyste surinfecté	2 (1,6)	2 (1,9)	0
Scarlatine	1 (0,8)	1 (1,0)	0
Infection diverse	17 (13,9)	17 (13,9)	0
Pleuro-pneumopathie	8 (6,6)	8 (7,6)	0
Pyélonéphrite aiguë	2 (1,6)	2 (1,9)	0
Empyème extradural	1 (0,8)	1 (1,0)	0
Empyème sous-dural	1 (0,8)	1 (1,0)	0
Méningite	1 (0,8)	1 (1,0)	0
Diverticulite	1 (0,8)	1 (1,0)	0
Iléite infectieuse	1 (0,8)	1 (1,0)	0
Endocardite infectieuse avec abcès multiples	1 (0,8)	1 (1,0)	0
Bactériémie	1 (0,8)	1 (1,0)	0
Prise en charge			
Antibiothérapie précédent l'admission, n (%)	44 (36,1)	33 (31,4)	11 (64,7)
Prescription d'une antibiothérapie, n (%) (N=121)	111 (91,7)	94 (90,4)	17 (100,0)
Réalisation d'un drainage chirurgical, n (%) ^f	23 (18,9)	17 (27,0)	6 (37,5)
Identification de l'agent pathogène, n (%) ^g	14 (11,5)	12 (11,4)	2 (11,8)
Mention arrêt/CI à l'AINS, n (%) ^h	33 (27,0)	28 (26,7)	5 (29,4)
Evolution, n (%)			
Retour à domicile	91 (74,6)	76 (72,4)	15 (88,2)
Hospitalisation	31 (25,4)	29 (27,6)	2 (11,8)

AINS : anti-inflammatoire non stéroïdien ; ORL : oto-rhino-laryngologie ; CI : contre-indication ; IQR : intervalle interquartile.

^a Patients dont l'âge était supérieur ou égal à 15 ans et 3 mois.

^b Unité fonctionnelle dans laquelle était produit le document.

^c Temps écoulé entre la dernière prise d'AINS documentée et l'admission.

^d Non précisé, le cas échéant.

^e Temps écoulé entre le début rapporté des symptômes et l'admission.

^f Si pathologie indiquée (abcès, cellulite).

^g En cours ou non recherché, le cas échéant.

^h Un arrêt ou une contre-indication à l'AINS était explicitement mentionnée par le clinicien.

Aucune notification au CRPV du CHU de Lille, concernant l'une des pathologies infectieuses identifiées en lien avec une exposition à un AINS, n'était retrouvée sur la période d'étude.

Discussion

1 Principaux résultats

Cette étude exploratoire a permis de développer et d'évaluer une première méthode de détection automatisée de cas d'infections bactériennes associées à une prise d'AINS à partir des données de l'EDS.

La requête a permis de réduire considérablement le volume de données à analyser : à partir d'un corpus initial d'environ 100 000 documents exploitables sur une période de 4 mois, 1 198 (1,2%) documents étaient pertinents pour analyse.

Sur un total de 769 documents analysés, l'annotation experte a permis de détecter 122 (15,9%) cas éligibles à une analyse d'imputabilité. La concordance inter-expert, avec des coefficients kappa de 75,0% et 85,9%, respectivement sur les données de 2018 et 2023.

Sur les 122 cas éligibles, 68 (55,7 %) étaient de sexe masculin. La majorité étaient des adultes (n=107, 87,7%) avec un âge médian (IQR) de 36,7 ans (25,7—49,6). Les passages aux urgences concernaient 109 (89,3 %) patients. Les complications infectieuses les plus fréquemment diagnostiquées étaient des cas de cellulites dentaires (n=26, 21,3 %), des abcès dentaires (n=17, 13,9 %) et des angines bactériennes (n=13, 10,7 %).

Concernant la prise d'AINS, 68 (55,7 %) patients déclaraient avoir consommé de l'ibuprofène préalablement à l'admission, tandis que 16 (13,1 %) déclaraient avoir consommé une autre molécule, dont le kétoprofène (n=7, 5,7%) et le diclofénac (n=6, 4,9%). La mention d'une prise d'AINS sans précision était renseignée pour 38 (31,1 %) patients. Une prescription médicale était déclarée pour 57 (54,3 %) patients, tandis qu'une automédication était déclarée chez 48 (45,7%) patients. L'arrêt ou la contre-indication de l'AINS à la suite de la prise en charge était explicitement mentionné par le clinicien pour 33 (27,0 %) patients. Le retour à domicile était indiqué pour 91 (74,6 %) patients, tandis que 31 (25,4 %) étaient transférés en hospitalisation. La durée médiane de séjour (IQR) était de 13,8 heures (9,2—30,3).

Sur l'ensemble de test (documents annotés sur la période 2023), l'annotation automatisée réalisée par le LLM, confrontée à l'annotation experte utilisée comme référence, rapportait un score F1 de 71,4% (Se 88,2%, VPP 60,0%).

2 Discussion des résultats

Parmi l'ensemble des cas jugés éligibles à une analyse d'imputabilité, le profil majoritaire était représenté par un adulte jeune autour de 35 ans, homme ou femme, sans antécédent notable, adressé aux urgences par le médecin traitant dans les 3 jours suivant la survenue de symptômes d'évolution défavorable. De façon plus ou moins concomitante, une prise ponctuelle ou répétée d'AINS, le plus souvent de l'ibuprofène, était associée. La prise pouvait être sur le mode de la prescription médicale, ou de l'automédication. Les pathologies les plus fréquemment décrites étaient la survenue ou l'aggravation d'une infection dentaire ou des tissus mous environnants à type de cellulite ou d'abcès, pour laquelle un drainage chirurgical

était parfois réalisé. Le retour à domicile était indiqué avec prescription d'une antibiothérapie (amoxicilline/acide clavulanique le plus souvent), d'un traitement antalgique approprié (paracétamol + palier II le plus souvent), ainsi qu'un suivi à 48h par médecin traitant ou le dentiste, ou une consigne de reconsulter en l'absence d'amélioration.

Les caractéristiques démographiques de notre population, adultes jeunes et enfants d'âge scolaire, sans prédominance de sexe marquée, correspondent aux populations à risque décrites dans le rapport actualisé de 2024 des CRPV de Tours et Marseille¹⁶. Celle-ci relevait que 21% des effets indésirables graves déclarés avec l'ibuprofène concernaient la survenue ou l'aggravation d'infection bactérienne touchant principalement des sujets jeunes sans facteur de risque particulier.

La prédominance des infections bucco-dentaires retrouvée dans notre étude (35,2% avec cellulites et abcès dentaires) était également attendue. Une étude prospective menée au CHU de Lille sur 267 patients avait déjà exploré cette association spécifique, bien que sans démontrer un lien de causalité entre prise d'AINS et aggravation des cellulites dentaires⁶¹. Néanmoins, la plupart des stomatologues et chirurgiens maxillo-faciaux constataient dans leur pratique quotidienne que l'exposition à un AINS pouvait être associée à la survenue de ces infections.

L'ibuprofène, AINS le plus consommé en France et impliqué dans plus de la majorité des cas retenus, confirme sa position déjà documentée de premier AINS en cause dans les complications infectieuses, essentiellement à streptocoque ou à pneumocoque. Celles-ci ont été observées après de très courtes durées de traitement (2 à 3 jours), y compris lorsque la prise d'AINS était associée à une antibiothérapie²³. Cela concorde avec notre observation d'un délai médian inférieur à 24 heures entre la dernière prise et l'admission, parmi 50% des cas pour lesquels l'information était renseignée, et la présence d'une antibiothérapie préalable chez plus d'un tiers des patients, n'ayant pas empêché l'évolution défavorable.

La proportion d'automédication déclarée (47%) reflète une problématique déjà documentée. Une étude menée dans la Somme en 2018 retrouvait que 67% des patients pratiquaient l'automédication, le symptôme le plus fréquent motivant le recours à l'automédication étant les céphalées (68,6%)¹². Cette pratique largement répandue, associée à la disponibilité des AINS sans ordonnance malgré leur retrait du libre accès en 2020, contribue au risque de complications infectieuses en retardant le diagnostic et la prise en charge adaptée.

La mention explicite d'un arrêt ou d'une contre-indication à l'AINS par le clinicien était présente dans 27% des cas. Cela pouvait témoigner d'une certaine sensibilisation au risque infectieux lié aux AINS, toutefois encore insuffisante. Cette observation contraste avec les recommandations répétées de l'ANSM depuis 2019, préconisant l'utilisation du paracétamol en première intention pour le traitement de la douleur et de la fièvre, particulièrement en contexte infectieux. La prescription ou le maintien d'une antibiothérapie quasi-systématique sans arrêt concomitant de l'AINS peut suggérer une méconnaissance persistante du risque.

3 Discussion de la méthode

3.1 Forces

L'implémentation d'un LLM pour l'annotation automatisée a montré des performances acceptables pour l'optimisation d'un processus d'amplification de signal en pharmacovigilance, dans le cadre du risque infectieux associé à la prise d'AINS. La méthodologie développée présente un caractère généralisable et transposable à d'autres problématiques nécessitant l'analyse de données textuelles massives. La stratégie de *prompt engineering* mise en œuvre peut constituer la base d'un cadre méthodologique reproductible.

La sensibilité observée reflète un compromis méthodologique assumé, privilégiant la réduction du temps d'analyse plutôt qu'une exhaustivité absolue jugée peu compatible avec les contraintes de la surveillance en routine. L'absence de cas effectivement déclarés au CRPV sur la période d'étude montre l'apport quantitatif de la stratégie proposée. Celle-ci pourrait permettre, occasionnellement, de contourner les limites structurelles de la notification spontanée et réduire le phénomène de sous-notification, pour pouvoir confirmer un signal.

La spécificité observée confirme la capacité du LLM à écarter efficacement les cas non pertinents, réduisant ainsi la charge de travail liée à l'analyse inutile de faux positifs. Bien qu'elle reste perfectible, cette méthode de pré-filtrage demeure particulièrement adaptée aux contraintes opérationnelles de la pharmacovigilance, où les ressources sont limitées par volume données à analyser et la gestion des tâches de routine en parallèle.

Le rendement observé avec l'annotation automatisée représente une amélioration en comparaison à une revue experte exhaustive : le LLM permet de concentrer l'expertise sur un corpus préfiltré de texte qu'on suppose davantage susceptible de contenir des cas intérêt en comparaison à un échantillon tiré aléatoirement dans l'EDS. Sur la base des performances observées, cette approche pourrait réduire d'environ 80% le volume de documents nécessitant une revue experte, tout en préservant 90% des cas réellement éligibles.

3.2 Limites

La qualité variable des données textuelles constitue une limite intrinsèque de la stratégie proposée. C'est particulièrement vrai en ce qui concerne les observations médicales issues des urgences, où la qualité informationnelle est tributaire à la fois d'une forte variabilité individuelle et des contraintes liées à l'activité. Les textes médicaux présentent fréquemment des informations incomplètes ou fragmentées, avec des formulations ambiguës nécessitant parfois une interprétation contextuelle. Cette hétérogénéité, notamment par l'absence fréquente de datation précise du début des symptômes et de la prise médicamenteuse, peut compromettre l'analyse fine de la chronologie et l'évaluation des critères d'imputabilité. L'information relative à la consommation d'AINS peut être sujette à une sous-déclaration ou un sous-recueil : d'une part les patients n'évoquent pas systématiquement leur consommation, d'autre part les cliniciens peuvent omettre de documenter cette exposition, introduisant un biais de mesure.

Il convient de souligner le caractère exploratoire de notre approche, les cas identifiés et décrits n'ayant pas fait l'objet d'une analyse d'imputabilité. Cette étape préliminaire de description de signal pourrait être complétée par une approche pharmaco-épidémiologique au sein même de

l'EDS. Par exemple, la réalisation d'une étude cas-témoins, comparant des cas avec infection documentée à des témoins sans infection, appariés sur l'âge, le sexe, la période et le motif de recours aux urgences, permettrait de mesurer l'association entre une exposition à un AINS et la survenue de complications infectieuses. La conduite d'une étude épidémiologique comparative apporterait un niveau de preuve supérieur à celui de l'analyse descriptive exploratoire réalisée pour les cas éligibles dans ce travail.

Bien qu'exhaustive, la restriction de l'analyse à une population hospitalière questionne la représentativité de nos résultats vis-à-vis de l'ensemble des complications infectieuses associées à la prise d'AINS en population générale. Les cas identifiés sont susceptibles de correspondre à des formes plus sévères d'infections nécessitant une venue à l'hôpital, excluant de fait les épisodes bénins traités en médecine de ville. La complémentarité avec le SNDS, recommandée par la HAS via des appariements systématiques EDS-SNDS participe à réduire ce biais en complétant la richesse clinique hospitalière avec le parcours de soin ambulatoire. Cette approche reste encore difficile à mettre en œuvre notamment par contraintes réglementaires, sans compter qu'elle demeure limitée dans notre contexte en raison de la problématique liée à l'automédication.

La période d'étude restreinte de deux mois sur chaque année (janvier et juillet) pourrait être influencée par des variations saisonnières et ne couvre qu'une fraction limitée de l'incidence annuelle.

Le ciblage a priori de certains services hospitaliers pouvait orienter la recherche vers des profils cliniques attendus et limiter la découverte de nouveaux signaux. Nous avons observé dans cette étude que la plupart des cas retenus concernait des passages aux urgences avec prise en charge rapide et retour au domicile. La recherche dans les autres services indépendamment du type de document à analyser (lettre de sortie, compte-rendu d'hospitalisation ou de consultation) s'est avérée peu pertinente pour retrouver des cas. Une exploration complémentaire pourrait consister en la démarche inverse : rechercher l'information dans un type particulier de document que sont les observations médicales, indépendamment du service d'origine (hors urgences). Sans viser l'exhaustivité, cette approche pourrait permettre l'identification d'autres profils cliniques. Leur extraction, désormais réalisable, ne l'était pas au moment de cette étude.

De façon similaire, notre stratégie de ciblage des typologies infectieuses, qui s'appuyait sur les signaux identifiés dans les précédentes enquêtes nationales de pharmacovigilance, constituait une approche plutôt conservatrice. Des analyses complémentaires pourraient être menées avec un périmètre d'entités pathologiques moins restrictif, permettant d'explorer des associations moins documentées.

L'approche exploratoire présentée dans cette étude, recherchant simultanément de multiples associations entre différents AINS et diverses pathologies infectieuses, contraste avec la notification spontanée qui documente un couple médicament-effet et un lien supposé entre les deux. Cette complexité méthodologique, nécessaire pour capturer la diversité des profils cliniques, constitue un cas d'usage particulièrement exigeant : les résultats présentés ici pourraient ne pas être représentatifs de ceux observés pour des problématiques plus circonscrites.

Bien qu'efficace pour analyser et extraire de l'information en contexte médical, le LLM ne dispose pas de réelle expertise pour interpréter les situations cliniques complexes ou ambiguës. Cela compromet la fiabilité de l'extraction dans les cas atypiques et limitant la

reproductibilité de la classification lorsque les critères d'éligibilité ne sont pas explicitement mentionnés dans le texte. La complexité d'interprétation pouvant exister chez le lecteur humain constitue *a fortiori* une limite pour le LLM. Le problème lié à la nature hétérogène de la donnée textuelle, comme évoquée précédemment, vient se surajouter à cette limitation.

Dans le cadre de cette étude, nous avons utilisé un LLM pré-entraîné, sans adaptation spécifique aux textes médicaux en langue française, en adoptant une approche basée uniquement sur le développement et l'optimisation itérative des instructions. Cette stratégie présente plusieurs avantages opérationnels, notamment une implémentation rapide sans nécessiter de corpus d'entraînement annoté, et une flexibilité permettant l'adaptation immédiate des critères de classification par simple modification du prompt. Ces avantages se font toutefois au prix de performances potentiellement suboptimales par rapport à des approches plus sophistiquées, l'absence de spécialisation pouvant limiter sa compréhension des nuances cliniques et de la terminologie française. Des analyses complémentaires pourraient inclure l'entraînement du modèle sur des textes médicaux français (*fine-tuning*), l'utilisation de modèles spécifiquement entraînés sur des données biomédicales, ou l'implémentation d'approches permettant au modèle d'apprendre à partir d'un nombre limité d'exemples annotés.

La dépendance au *prompt engineering*, bien que flexible, introduisait une variabilité dans les résultats selon la formulation des instructions. Plusieurs explorations complémentaires (non présentées ici) ont été réalisées, notamment une analyse qualitative de réponses textuelles générées par le LLM, ou encore une analyse multivariée visant à déterminer l'impact de variations contrôlées dans la structure du prompt.

L'étude réalisée consistait à détecter des cas éligibles à une analyse détaillée par des experts pharmacologues selon des critères d'imputabilité. Cette approche de détection et de description de signal constituait une étape préliminaire et exploratoire, dont les résultats devront être confirmés, tout d'abord, par une description exploratoire sur les cas confirmés par l'analyse d'imputabilité et, ensuite, par la conduite d'études épidémiologiques analytiques offrant un meilleur niveau de preuve.

4 Perspectives

Cette étude préliminaire et exploratoire a pu montrer la capacité de notre méthode à extraire, à partir de l'EDS, des cas éligibles à une analyse détaillée en pharmacovigilance. Les cas identifiés devraient faire l'objet d'une expertise d'imputabilité en vue d'une éventuelle déclaration à l'ANSM. Au-delà, l'EDS permettrait de conduire des études pharmaco-épidémiologiques, notamment cas-témoins, afin de quantifier plus formellement l'association entre AINS et complications infectieuses.

Sous réserve d'analyses complémentaires et d'une évaluation de la faisabilité de sa mise en pratique, la stratégie de détection proposée pourrait à terme contribuer à une surveillance plus systématique et exhaustive du risque infectieux lié aux AINS, et éventuellement ouvrir le champ vers d'autres problématiques.

La disponibilité quasi-continue des données dans l'EDS est un atout pour la détection active et la réactivité face à de potentiels signaux émergents. Nous pourrions envisager à terme le déploiement d'un système de détection automatisée sur un signal particulier analysant

l'ensemble des textes médicaux produites en fil de l'eau dans des secteurs stratégiques. A la manière du système de notification spontanée déjà en place en pharmacovigilance, un système d'alerte automatisé pourrait parallèlement notifier, de façon quotidienne ou hebdomadaire, la détection de cas jugés éligibles à une analyse approfondie à partir de l'EDS.

Les résultats de cette étude évoquent la nécessité d'action pour prévenir les complications infectieuses potentiellement associées aux AINS. La sensibilisation des prescripteurs au risque infectieux apparaît prioritaire, notamment en médecine de ville où persistent des co-prescriptions inappropriées d'AINS et d'antibiotiques. Cette sensibilisation devrait s'accompagner d'une réévaluation du rapport bénéfice-risque des AINS dans les contextes infectieux, privilégiant le paracétamol en première intention conformément aux recommandations actuelles. Face à la part d'automédication observée, un avertissement sur les boîtes d'AINS concernant le risque infectieux en cas de fièvre pourrait être envisagé, sur le modèle de celui existant pour le paracétamol concernant le risque de surdosage. Le renforcement de la surveillance clinique en cas d'infection débutante sous AINS permettrait une détection précoce des complications, tandis que l'information des patients sur les signes d'alerte favoriserait une prise en charge plus précoce.

Au niveau réglementaire, l'identification systématique de cas documentés par la méthode proposée, sous réserves d'études complémentaires, pourrait participer à fournir des arguments supplémentaires aux instances européennes visant à renforcer la position sur le risque infectieux associé aux AINS, justifiant une actualisation des recommandations.

Liste des tables

Table 1. Descriptif de la population retenue pour annotation, globalement et selon l'année.	33
Table 2. Table de contingence croisant les annotations de chaque expert avant consensus, année 2018.....	34
Table 3. Table de contingence croisant les annotations de chaque expert avant consensus, année 2023.....	34
Table 4. Table de contingence croisant l'annotation experte avec l'annotation automatisée (classification déclarée), année 2018.	35
Table 5. Table de contingence croisant l'annotation experte avec l'annotation automatisée (classification théorique), année 2018.	35
Table 6. Table de contingence croisant l'annotation experte avec l'annotation automatisée (classification déclarée), année 2023.	36
Table 7. Descriptif des cas classés éligibles par l'annotation experte après consensus, globalement et selon l'année.	37

Liste des figures

Figure 1. Exemple de recherche d'expressions régulières couvrant le concept « érysipèle ».....	22
Figure 2. Groupe de concepts utilisés pour la requête textuelle.	24
Figure 3. Requête EDS et requête textuelle pour constitution d'un corpus de documents à annoter.	25
Figure 4. Capture écran de l'interface mise à disposition pour l'annotation experte (classification des cas).	27
Figure 5. Diagramme de flux.	32
Figure 6. Nombre de cas classés éligibles par l'annotation experte après consensus, à partir de la population retenue pour annotation, sur l'ensemble de période d'étude.	33

Références

1. Anti-inflammatoires non-stéroïdiens [Internet]. [cited 2025 Sep 1];Available from: <https://pharmacomedicale.org/medicaments/par-specialites/item/anti-inflammatoires-non-steroidiens>
2. Ricciotti E, FitzGerald GA. Prostaglandins and inflammation. *Arterioscler Thromb Vasc Biol* 2011;31(5):986–1000.
3. Prajitha N, Ss A, Pv M. Comprehensive biology of antipyretic pathways. *Cytokine* 2019;116:120–7.
4. Hider-Mlynarz K, Cavalié P, Maison P. Trends in analgesic consumption in France over the last 10 years and comparison of patterns across Europe. *Br J Clin Pharmacol* 2018;84(6):1324–34.
5. Malbos D. Ibuprofène et paracétamol, promouvoir le bon usage. *Actual Pharm* 2020;59(598):20–2.
6. Paitraud D. Kétoprofène sous forme topique et risque de photosensibilisation : rappel des règles de bon usage [Internet]. VIDAL. 2014 [cited 2025 Sep 1];Available from: <https://www.vidal.fr/actualites/14018-ketoprofene-sous-forme-topique-et-risque-de-photosensibilisation-rappel-des-regles-de-bon-usage.html>
7. Schmidt M, Sørensen HT, Pedersen L. Diclofenac use and cardiovascular risks: series of nationwide cohort studies. *BMJ* 2018;362:k3426.
8. McGettigan P, Henry D. Cardiovascular risk with non-steroidal anti-inflammatory drugs: systematic review of population-based controlled observational studies. *PLoS Med* 2011;8(9):e1001098.
9. Micallef J, Soeiro T, Jonville-Béra A-P. Non-steroidal anti-inflammatory drugs, pharmacology, and COVID-19 infection. *Therapie* 2020;75(4):355–62.
10. Duong M, Gulmez SE, Salvo F, et al. Usage patterns of paracetamol in France. *Br J Clin Pharmacol* 2016;82(2):498–503.
11. Le Turnier P, Bouteille D, Joyau C, Veyrac G, Afferay N. Bacterial infections and NSAIDs exposure? Seek septic complications. *Eur J Intern Med* 2017;41:e33–4.
12. Anger V. Automédication par anti-inflammatoires non stéroïdiens (AINS) : état des lieux des connaissances et des pratiques de patients dans la Somme en 2018. 2019;67.
13. Composantes du SNDS [Internet]. [cited 2025 Sep 2];Available from: <https://www.snds.gouv.fr/SNDS/Composantes-du-SNDS>
14. La déclaration des effets indésirables [Internet]. Ministère Trav. Santé Solidar. Fam. [cited 2025 Sep 30];Available from: <https://sante.gouv.fr/soins-et-maladies/medicaments/la-surveillance-des-medicaments/article/la-declaration-des-effets-indesirables>
15. Base de Données Publique des Médicaments [Internet]. [cited 2025 Aug 28];Available from: <https://base-donnees-publique.medicaments.gouv.fr/medicament/61788736/extrait#tab-rcp>
16. Comité scientifique permanent Pharmaco-surveillance et bon usage - Formation restreinte expertise et bon usage [Internet]. 2024 [cited 2024 Nov 24]. Available from: <https://ansm.sante.fr/evenements/comite-scientifique-permanent-pharmaco-surveillance-et-bon-usage-formation-restreinte-expertise-et-bon-usage-2>
17. Stevens DL, Bryant AE. Necrotizing Soft-Tissue Infections. *N Engl J Med* 2017;377(23):2253–65.
18. Agard M, Asakrah S, Morici LA. PGE2 suppression of innate immunity during mucosal bacterial infection. *Front Cell Infect Microbiol* [Internet] 2013 [cited 2025 Sep 8];3. Available from: <https://www.frontiersin.org/journals/cellular-and-infection-microbiology/articles/10.3389/fcimb.2013.00045/full>
19. Kalinski P. Regulation of Immune Responses by Prostaglandin E2. *J Immunol Baltim Md* 1950 2012;188(1):21–8.
20. Khayrullina T, Yen J-H, Jing H, Ganea D. In vitro differentiation of dendritic cells in the presence of prostaglandin E2 alters the IL-12/IL-23 balance and promotes differentiation of Th17 cells. *J Immunol Baltim Md* 1950 2008;181(1):721–35.
21. Boniface K, Bak-Jensen KS, Li Y, et al. Prostaglandin E2 regulates Th17 cell differentiation and function through cyclic AMP and EP2/EP4 receptor signaling. *J Exp Med* 2009;206(3):535–48.
22. Bryant AE, Bayer CR, Aldape MJ, Stevens DL. The roles of injury and nonsteroidal anti-inflammatory drugs in the development and outcomes of severe group A streptococcal soft tissue infections. *Curr Opin Infect Dis* 2015;28(3):231–9.

23. Comité technique de pharmacovigilance - Infections bactériennes graves rapportées avec l'ibuprofène ou le kétoprofène dans le traitement symptomatique de la fièvre ou de douleur non rhumatologique [Internet]. 2019 [cited 2025 Sep 2]. Available from: <https://ansm.sante.fr/uploads/2020/10/29/rapport-pv-ains-tours-marseille-2019.pdf>
24. Anti-inflammatoires non stéroïdiens (AINS) et complications infectieuses graves [Internet]. ANSM. [cited 2025 Aug 28];Available from: <https://ansm.sante.fr/actualites/anti-inflammatoires-non-steroidiens-ains-et-complications-infectieuses-graves>
25. L'utilisation d'anti-inflammatoires non stéroïdiens (AINS), dans le traitement de la fièvre et/ou de la douleur, n'est pas recommandée chez l'enfant atteint de varicelle - ANSM [Internet]. [cited 2025 Aug 28];Available from: <https://archive.anmsante.fr/S-informer/Informations-de-securite-Lettres-aux-professionnels-de-sante/L-utilisation-d-anti-inflammatoires-non-steroidiens-AINS-dans-le-traitement-de-la-fievre-et-ou-de-la-douleur-n-est-pas-recommandee-chez-l-enfant-atteint-de-varicelle>
26. Bon usage du paracétamol et des anti-inflammatoires non stéroïdiens (AINS) : l'ANSM veut renforcer le rôle de conseil du pharmacien [Internet]. ANSM. [cited 2025 Aug 28];Available from: <https://ansm.sante.fr/actualites/bon-usage-du-paracetamol-et-des-anti-inflammatoires-non-steroidiens-ains-lansm-veut-renforcer-le-role-de-conseil-du-pharmacien>
27. Recommendations on Safety Signals [Internet]. [cited 2025 Sep 1]. Available from: https://www.ema.europa.eu/en/documents/prac-recommendation/prac-recommendations-signals-adopted-14-17-april-2020-prac-meeting_en.pdf
28. Recommendations on Safety Signals | European Medicines Agency (EMA) [Internet]. 2013 [cited 2025 Aug 31];Available from: <https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/pharmacovigilance-post-authorisation/signal-management/prac-recommendations-safety-signals>
29. EMA gives advice on the use of non-steroidal anti-inflammatories for COVID-19 [Internet]. EMA. 2020 [cited 2025 Sep 1];Available from: <https://www.ema.europa.eu/en/news/ema-gives-advice-use-non-steroidal-anti-inflammatories-covid-19>
30. Missions de la pharmacovigilance [Internet]. RFCRPV. [cited 2025 Aug 31];Available from: <https://www.rfcrpv.fr/missions/>
31. Leca É. La pharmacovigilance : qui fait quoi ? [Internet]. VIDAL. 2023 [cited 2025 Aug 31];Available from: <https://www.vidal.fr/actualites/30392-la-pharmacovigilance-qui-fait-quoi.html>
32. Arimone Y, Bidault I, Dutertre J-P, et al. Réactualisation de la méthode française d'imputabilité des effets indésirables des médicaments. Thérapie 2011;66(6):517–25.
33. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions : a systematic review. Drug Saf 2006;29(5):385–96.
34. Pharmacovigilance: Overview | European Medicines Agency (EMA) [Internet]. 2017 [cited 2025 Aug 31];Available from: <https://www.ema.europa.eu/en/human-regulatory-overview/pharmacovigilance-overview>
35. Pharmacovigilance - APHP DAJDP [Internet]. [cited 2025 Aug 31];Available from: <https://affairesjuridiques.aphp.fr/textes/pharmacovigilance/>
36. Jonville-Bera AP, Gautier S, Micallef J, et al. Pharmacovigilance des médicaments et des vaccins contre la COVID-19 durant la pandémie : comment le Réseau français des centres régionaux de pharmacovigilance a relevé le défi ? Therapies 2023;78(5):477–88.
37. Durrieu G, Jacquot J, Mège M, et al. Completeness of Spontaneous Adverse Drug Reaction Reports Sent by General Practitioners to a Regional Pharmacovigilance Centre: A Descriptive Study. Drug Saf 2016;39(12):1189–95.
38. Lopez-Gonzalez E, Herdeiro MT, Figueiras A. Determinants of under-reporting of adverse drug reactions: a systematic review. Drug Saf 2009;32(1):19–31.
39. Humbert X, Jacquot J, Alexandre J, et al. Informativité des déclarations de pharmacovigilance en médecine générale en France. Santé Publique 2019;31(4):561–6.
40. Fusaroli M, Salvo F, Khouri C, Raschi E. The reporting of disproportionality analysis in pharmacovigilance: spotlight on the READUS-PV guideline. Front Pharmacol 2024;15:1488725.
41. Cutroneo PM, Sartori D, Tuccori M, et al. Conducting and interpreting disproportionality analyses derived from spontaneous reporting systems. Front Drug Saf Regul [Internet] 2024 [cited 2025 Aug 31];3. Available from: <https://www.frontiersin.org/journals/drug-safety-and-regulation/articles/10.3389/fdsfr.2023.1323057/full>
42. Caster O, Aoki Y, Gattepaille LM, Grundmark B. Disproportionality Analysis for Pharmacovigilance Signal Detection in Small Databases or Subsets: Recommendations for Limiting False-Positive Associations. Drug Saf 2020;43(5):479–87.
43. Naranjo CA, Busto U, Sellers EM, et al. A method for estimating the probability of adverse drug reactions. Clin Pharmacol Ther 1981;30(2):239–45.

44. Eiden C, Peyrière H. Comparaison de deux méthodes d'imputabilité des effets indésirables du voriconazole notifiés dans la base nationale de pharmacovigilance : Bégaud versus Naranjo. *Pharm Hosp* 2009;44(4):186–9.
45. Faillie J-L, Montastruc F, Montastruc J-L, Pariente A. L'apport de la pharmaco-épidémiologie à la pharmacovigilance. *Therapies* 2016;71(2):203–9.
46. Entrepôts de données de santé hospitaliers en France - Quel potentiel pour la Haute Autorité de santé ? [Internet]. [cited 2025 Sep 15];Available from: https://www.has-sante.fr/upload/docs/application/pdf/2022-11/rapport_entrepots_donnees_sante_hospitaliers.pdf
47. Explorez la cartographie des entrepôts de données de santé en France [Internet]. [cited 2025 Sep 15];Available from: <https://www.cnil.fr/fr/explorez-la-cartographie-des-entrepot-de-donnees-de-sante-en-france>
48. Fiches descriptives des standards et terminologies nationaux et internationaux des données de santé | Documentation du SNDS & SNDS OMOP [Internet]. [cited 2025 Sep 15];Available from: <https://www.documentation-snds.health-data-hub.fr/standards/>
49. Sedlakova J, Daniore P, Witsch AH, et al. Challenges and best practices for digital unstructured data enrichment in health research: A systematic narrative review. *PLOS Digit Health* 2023;2(10):e0000347.
50. El Aridi L, Jantzem H, Happe A, Cauvin JM, Carlhant-Kowalski D, Gourier G. HOPIPRAC : Interface de pharmacovigilance pour la détection de cas d'événements indésirables dans un entrepôt de données hospitalier. *Therapies* [Internet] 2025 [cited 2025 Sep 21];Available from: <https://www.sciencedirect.com/science/article/pii/S0040595725000277>
51. Raiaan MAK, Mukta MdSH, Fatema K, et al. A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. *IEEE Access* 2024;12:26839–74.
52. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need [Internet]. 2023 [cited 2025 Sep 22];Available from: <http://arxiv.org/abs/1706.03762>
53. Naveed H, Khan AU, Qiu S, et al. A Comprehensive Overview of Large Language Models [Internet]. 2024 [cited 2025 Sep 22];Available from: <http://arxiv.org/abs/2307.06435>
54. Touvron H, Lavril T, Izacard G, et al. LLaMA: Open and Efficient Foundation Language Models [Internet]. 2023 [cited 2025 Sep 22];Available from: <http://arxiv.org/abs/2302.13971>
55. Maaz S, Palaganas JC, Palaganas G, Bajwa M. A guide to prompt design: foundations and applications for healthcare simulationists. *Front Med* 2025;11:1504532.
56. Rust P, Pfeiffer J, Vulic I, Ruder S, Gurevych I. How Good is Your Tokenizer? On the Monolingual Performance of Multilingual Language Models [Internet]. 2021 [cited 2025 Sep 22];Available from: <http://arxiv.org/abs/2012.15613>
57. Labrak Y, Bazoge A, Dufour R, et al. DrBERT: A Robust Pre-trained Model in French for Biomedical and Clinical domains [Internet]. In: Rogers A, Boyd-Graber J, Okazaki N, editors. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Toronto, Canada: Association for Computational Linguistics; 2023 [cited 2025 May 4]. p. 16207–21. Available from: <https://aclanthology.org/2023.acl-long.896>
58. Touchent R, Romary L, De La Clergerie E. CamemBERT-bio: Leveraging Continual Pre-training for Cost-Effective Models on French Biomedical Data [Internet]. In: LREC-COLING 2024 - The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation. Torino, Italy: 2024 [cited 2025 Sep 26]. Available from: <https://hal.science/hal-04528508>
59. DELIBERATION n°2019-103 du 5 SEPTEMBRE 2019 [Internet]. [cited 2025 Sep 23]. Available from: <https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000039292712/>
60. Llama3.3 - New state of the art 70B model. Llama 3.3 70B offers similar performance compared to the Llama 3.1 405B model. [Internet]. [cited 2025 Sep 22];Available from: <https://ollama.com/llama3.3>
61. Nicot R, Hippy C, Hochart C, et al. Les anti-inflammatoires aggravent-ils les cellulites faciales d'origine dentaire ? *Rev Stomatol Chir Maxillo-Faciale Chir Orale* 2013;114(5):304–9.

Annexes

Annexe 1. Modèle de prompt « 0 » (A0_B0_C0_D0_E0).

IDENTITÉ

Tu es un système d'extraction d'informations médicales spécialisé en pharmacovigilance.

OBJECTIF

À partir du texte médical, identifie les cas d'infections bactériennes possiblement associées à une exposition préalable à un anti-inflammatoire non stéroïdien (AINS).

– DÉBUT DU TEXTE MÉDICAL – {input} – FIN DU TEXTE MÉDICAL –

CRITÈRES D'ÉLIGIBILITÉ

1. L'infection bactérienne ou les symptômes précurseurs doivent constituer le motif de venue à l'hôpital.

2. La nature bactérienne de l'infection doit être confirmée par la présence d'au moins 1 critère parmi les 3 suivants :

- Prescription d'une antibiothérapie
- Réalisation d'un drainage chirurgical
- Identification de l'agent pathogène

3. La dernière exposition à un AINS doit avoir eu lieu dans les jours précédant la venue à l'hôpital.

RAPPELS

Voici les acronymes que tu dois connaître :

- ains : anti-inflammatoire non stéroïdien
- atb : antibiotique
- atg : antalgique

Voici une liste NON EXHAUSTIVE des éléments que tu dois identifier :

- infection bactérienne : cellulite, empyème, abcès, angine, otite, sinusite, érysipèle, dermohypodermite, conjonctivite, fasciite
- AINS : ibuprofène, kétoprofène, diclofénac, naproxène, coxibs, oxicams, nurofen, advil, antarène, spifen, profenid, buprofenid, voltarène, AINS, anti-inflammatoire sans précision

Les entités suivantes NE SONT PAS des infections bactériennes :

- syndrome grippal, syndrome pseudo-grippal
- rhinopharyngite ou pharyngite sans précision

Les médicaments suivants NE SONT PAS des AINS :

- paracétamol, doliprane
- tramadol, codéine

- aspirine, aspégic, kardégic
- corticoïdes

RÉPONSE ATTENDUE

Tu dois produire une réponse au format JSON strictement structurée comme suit :

```
```json
{{

 "presence_infection": boolean (le texte fait-il mention d'une admission à l'hôpital
 pour une infection en cours ou des symptômes précurseurs d'infection ?),

 "infection_bactérienne": boolean (la nature bactérienne de l'infection est-elle
 confirmée ?),

 "exposition_ains": boolean (le texte fait-il mention d'une exposition à un AINS
 préalablement à la venue à l'hôpital ?),

 "cas_eligible": boolean (sur la base de tes réponses, détermine si le cas répond aux
 critères d'éligibilité)

}}
```
- Le texte doit obligatoirement faire mention d'une infection bactérienne en cours ou de
symptômes précurseurs ayant motivé la venue à l'hôpital.

- Ta conclusion ne doit pas porter uniquement sur l'interprétation d'un bilan biologique.

- Si l'infection est uniquement mentionnée dans les antécédents, tu ne dois pas la
considérer comme étant une infection en cours.

- La prise d'AINS doit impérativement précéder la survenue ou l'aggravation de l'infection.

- La dernière prise doit systématiquement être considérée comme étant inférieure à 3 jours
avant la venue à l'hôpital si l'AINS fait partie des traitements habituels.

- La mention d'une prise avant la venue à l'hôpital sans précision du délai doit t'inciter
à inclure le cas.

- La mention d'une prise uniquement durant le séjour à l'hôpital doit t'inciter à exclure
le cas.

- La mention d'une allergie à un AINS doit t'inciter à exclure le cas.

N'explicite PAS ton raisonnement.
```

Annexe 2. Modèle de prompt « 1 » (A1_B1_C1_D1_E1).

OBJECTIF

```
```json
{{

 "objectif": "à partir du texte médical, identifie les cas d'infections bactériennes
 possiblement associées à une exposition préalable à un anti-inflammatoire non
 stéroïdien (AINS)",

 "texte_médical": "{input}",

}}
````
```

CRITÈRES D'ÉLIGIBILITÉ

```
```json
{{

 "présence_infection": "l'infection bactérienne ou les symptômes précurseurs doivent
 constituer le motif de venue à l'hôpital",

 "infection_bactérienne": "la nature bactérienne de l'infection doit être confirmée par
 la présence d'au moins 1 critère parmi les 3 suivants : 1) prescription d'une
 antibiothérapie et/ou 2) réalisation d'un drainage chirurgical et/ou 3) identification
 de l'agent pathogène",

 "exposition_ains": "la dernière exposition à un AINS doit avoir eu lieu dans les jours
 précédant la venue à l'hôpital"

}}
````
```

RAPPELS

```
{{

  "acronymes": {{

    "ains": "anti-inflammatoire non stéroïdien",
    "atb": "antibiotique",
    "atg": "antalgique"
  }},

  "infection_bactérienne": {{

    "true": [
      "cellulite", "empyème", "abcès", "angine", "otite", "sinusite", "érysipèle",
      "dermohypodermite", "conjonctivite", "fasciite"
    ],
    "false": [
      ...
    ]
  }}
}}
```

```

    "syndrome grippal", "syndrome pseudo-grippal", "rhinopharyngite", "pharyngite
    sans précision"

    ],
}

"AINS" : {{

    "true": [
        "ibuprofène", "kétoprofène", "diclofénac", "naproxène", "coxibs", "oxicams",
        "nurofen", "advil", "antarène", "spifen", "profenid", "biprofenid",
        "voltarène", "AINS", "anti-inflammatoire sans précision"
    ],
    "false": [
        "paracétamol", "doliprane", "tramadol", "codéine", "aspirine", "aspégic",
        "kardégic", "corticoïdes"
    ]
}
}}
```

RÉPONSE ATTENDUE

Tu dois produire une réponse au format JSON strictement structurée comme suit :

```

```json
{{

 "presence_infection": boolean (le texte fait-il mention d'une admission à l'hôpital
 pour une infection en cours ou des symptômes précurseurs d'infection ?),

 "infection_bactérienne": boolean (la nature bactérienne de l'infection est-elle
 confirmée ?),

 "exposition_ains": boolean (le texte fait-il mention d'une exposition à un AINS
 préalablement à la venue à l'hôpital ? la dernière prise doit systématiquement être
 considérée comme étant inférieure à 3 jours avant la venue à l'hôpital si l'AINS fait
 partie des traitements habituels ; la mention d'une prise avant la venue à l'hôpital
 sans précision du délai doit t'inciter à inclure le cas ; la mention d'une prise
 uniquement durant le séjour à l'hôpital doit t'inciter à exclure le cas ; la mention
 d'une allergie à un AINS doit t'inciter à exclure le cas),

 "cas_eligible": boolean (sur la base de tes réponses, détermine si le cas répond aux
 critères d'éligibilité)

}}
```

```

N'explicite PAS ton raisonnement.

Annexe 3. Prompt sélectionné pour annotation sur l'ensemble de test (A0_B1_C0_D0_E0).

IDENTITÉ

Tu es un système d'extraction d'informations médicales spécialisé en pharmacovigilance.

OBJECTIF

```json

{ {

    "objectif": "à partir du texte médical, identifie les cas d'infections bactériennes possiblement associées à une exposition préalable à un anti-inflammatoire non stéroïdien (AINS)",

    "texte\_médical": "{input}",

}}

...

**### CRITÈRES D'ÉLIGIBILITÉ**

1. L'infection bactérienne ou les symptômes précurseurs doivent constituer le motif de venue à l'hôpital.
2. La nature bactérienne de l'infection doit être confirmée par la présence d'au moins 1 critère parmi les 3 suivants :
  - Prescription d'une antibiothérapie
  - Réalisation d'un drainage chirurgical
  - Identification de l'agent pathogène
3. La dernière exposition à un AINS doit avoir eu lieu dans les jours précédant la venue à l'hôpital.

**### RAPPELS**

Voici les acronymes que tu dois connaître :

- ains : anti-inflammatoire non stéroïdien
- atb : antibiotique
- atg : antalgique

Voici une liste NON EXHAUSTIVE des éléments que tu dois identifier :

- infection bactérienne : cellulite, empyème, abcès, angine, otite, sinusite, érysipèle, dermohypodermite, conjonctivite, fasciite
- AINS : ibuprofène, kétoprofène, diclofénac, naproxène, coxibs, oxicams, nurofen, advil, antarène, spifen, profenid, buprofenid, voltarène, AINS, anti-inflammatoire sans précision

Les entités suivantes NE SONT PAS des infections bactériennes :

- syndrome grippal, syndrome pseudo-grippal
- rhinopharyngite ou pharyngite sans précision

Les médicaments suivants NE SONT PAS des AINS :

- paracétamol, doliprane

- tramadol, codéine
- aspirine, aspégic, kardégic
- corticoïdes

### ### RÉPONSE ATTENDUE

Tu dois produire une réponse au format JSON strictement structurée comme suit :

```
```json
{{
  "presence_infection": boolean (le texte fait-il mention d'une admission à l'hôpital
  pour une infection en cours ou des symptômes précurseurs d'infection ?),
  "infection_bactérienne": boolean (la nature bactérienne de l'infection est-elle
  confirmée ?),
  "exposition_ains": boolean (le texte fait-il mention d'une exposition à un AINS
  préalablement à la venue à l'hôpital ?),
  "cas_eligible": boolean (sur la base de tes réponses, détermine si le cas répond aux
  critères d'éligibilité)
}}
```
- Le texte doit obligatoirement faire mention d'une infection bactérienne en cours ou de
symptômes précurseurs ayant motivé la venue à l'hôpital.
- Ta conclusion ne doit pas porter uniquement sur l'interprétation d'un bilan biologique.
- Si l'infection est uniquement mentionnée dans les antécédents, tu ne dois pas la
considérer comme étant une infection en cours.
- La prise d'AINS doit impérativement précéder la survenue ou l'aggravation de l'infection.
- La dernière prise doit systématiquement être considérée comme étant inférieure à 3 jours
avant la venue à l'hôpital si l'AINS fait partie des traitements habituels.
- La mention d'une prise avant la venue à l'hôpital sans précision du délai doit t'inciter
à inclure le cas.
- La mention d'une prise uniquement durant le séjour à l'hôpital doit t'inciter à exclure
le cas.
- La mention d'une allergie à un AINS doit t'inciter à exclure le cas.
N'explicite PAS ton raisonnement.
```

**Annexe 4 : Capacités discriminantes du LLM relatives aux 64 annotations automatisées réalisées sur l'ensemble d'apprentissage (classification déclarée).**

| n  | prompt         | spec_pm | confmat_vp | confmat_fn | confmat_fp | confmat_vn | confmat_total | metric_f1    | metric_se | metric_vpp |
|----|----------------|---------|------------|------------|------------|------------|---------------|--------------|-----------|------------|
| 1  | A0_B1_C0_D0_E0 | 0       | 92         | 13         | 38         | 483        | 626           | <b>0,783</b> | 0,876     | 0,708      |
| 2  | A0_B0_C0_D0_E0 | 1       | 93         | 12         | 41         | 480        | 626           | <b>0,778</b> | 0,886     | 0,694      |
| 3  | A1_B0_C1_D0_E0 | 1       | 92         | 13         | 40         | 481        | 626           | <b>0,776</b> | 0,876     | 0,697      |
| 4  | A0_B0_C0_D0_E0 | 0       | 93         | 12         | 42         | 479        | 626           | <b>0,775</b> | 0,886     | 0,689      |
| 5  | A1_B1_C1_D0_E0 | 0       | 91         | 14         | 39         | 482        | 626           | <b>0,774</b> | 0,867     | 0,700      |
| 6  | A0_B1_C1_D0_E1 | 1       | 91         | 14         | 40         | 481        | 626           | <b>0,771</b> | 0,867     | 0,695      |
| 7  | A1_B0_C0_D0_E0 | 0       | 94         | 11         | 45         | 476        | 626           | <b>0,770</b> | 0,895     | 0,676      |
| 8  | A0_B1_C0_D0_E0 | 1       | 92         | 13         | 42         | 479        | 626           | <b>0,770</b> | 0,876     | 0,687      |
| 9  | A1_B1_C1_D0_E1 | 0       | 90         | 15         | 39         | 482        | 626           | <b>0,769</b> | 0,857     | 0,698      |
| 10 | A1_B0_C1_D1_E1 | 0       | 95         | 10         | 47         | 474        | 626           | <b>0,769</b> | 0,905     | 0,669      |
| 11 | A0_B1_C1_D1_E1 | 1       | 93         | 12         | 44         | 477        | 626           | <b>0,769</b> | 0,886     | 0,679      |
| 12 | A1_B0_C0_D0_E1 | 1       | 94         | 11         | 46         | 475        | 626           | <b>0,767</b> | 0,895     | 0,671      |
| 13 | A0_B0_C1_D0_E1 | 0       | 90         | 15         | 40         | 481        | 626           | <b>0,766</b> | 0,857     | 0,692      |
| 14 | A1_B0_C1_D0_E0 | 0       | 90         | 15         | 40         | 481        | 626           | <b>0,766</b> | 0,857     | 0,692      |
| 15 | A1_B1_C1_D0_E0 | 1       | 90         | 15         | 40         | 481        | 626           | <b>0,766</b> | 0,857     | 0,692      |
| 16 | A1_B1_C0_D0_E0 | 0       | 93         | 12         | 45         | 476        | 626           | <b>0,765</b> | 0,886     | 0,674      |
| 17 | A1_B0_C0_D0_E1 | 0       | 96         | 9          | 50         | 471        | 626           | <b>0,765</b> | 0,914     | 0,658      |
| 18 | A1_B0_C1_D0_E1 | 1       | 91         | 14         | 42         | 479        | 626           | <b>0,765</b> | 0,867     | 0,684      |
| 19 | A0_B1_C1_D0_E0 | 1       | 89         | 16         | 39         | 482        | 626           | <b>0,764</b> | 0,848     | 0,695      |
| 20 | A1_B0_C0_D1_E0 | 0       | 97         | 8          | 52         | 469        | 626           | <b>0,764</b> | 0,924     | 0,651      |
| 21 | A0_B1_C1_D0_E0 | 0       | 90         | 15         | 41         | 480        | 626           | <b>0,763</b> | 0,857     | 0,687      |
| 22 | A0_B0_C1_D0_E1 | 1       | 90         | 15         | 41         | 480        | 626           | <b>0,763</b> | 0,857     | 0,687      |
| 23 | A0_B1_C1_D0_E1 | 0       | 88         | 17         | 38         | 483        | 626           | <b>0,762</b> | 0,838     | 0,698      |
| 24 | A1_B0_C0_D1_E0 | 1       | 96         | 9          | 51         | 470        | 626           | <b>0,762</b> | 0,914     | 0,653      |
| 25 | A1_B0_C0_D1_E1 | 1       | 96         | 9          | 51         | 470        | 626           | <b>0,762</b> | 0,914     | 0,653      |
| 26 | A0_B0_C1_D1_E1 | 0       | 91         | 14         | 43         | 478        | 626           | <b>0,762</b> | 0,867     | 0,679      |
| 27 | A1_B0_C1_D0_E1 | 0       | 91         | 14         | 43         | 478        | 626           | <b>0,762</b> | 0,867     | 0,679      |
| 28 | A1_B0_C0_D0_E1 | 1       | 94         | 11         | 48         | 473        | 626           | <b>0,761</b> | 0,895     | 0,662      |
| 29 | A0_B0_C1_D0_E0 | 1       | 89         | 16         | 40         | 481        | 626           | <b>0,761</b> | 0,848     | 0,690      |
| 30 | A1_B1_C0_D0_E0 | 1       | 92         | 13         | 45         | 476        | 626           | <b>0,760</b> | 0,876     | 0,672      |
| 31 | A1_B1_C1_D0_E1 | 1       | 92         | 13         | 45         | 476        | 626           | <b>0,760</b> | 0,876     | 0,672      |
| 32 | A0_B1_C1_D1_E1 | 0       | 93         | 12         | 47         | 474        | 626           | <b>0,759</b> | 0,886     | 0,664      |
| 33 | A0_B0_C1_D0_E0 | 0       | 88         | 17         | 39         | 482        | 626           | <b>0,759</b> | 0,838     | 0,693      |
| 34 | A1_B0_C1_D1_E1 | 1       | 94         | 11         | 49         | 472        | 626           | <b>0,758</b> | 0,895     | 0,657      |
| 35 | A0_B1_C0_D0_E1 | 0       | 93         | 12         | 48         | 473        | 626           | <b>0,756</b> | 0,886     | 0,660      |
| 36 | A0_B0_C0_D1_E1 | 1       | 93         | 12         | 48         | 473        | 626           | <b>0,756</b> | 0,886     | 0,660      |
| 37 | A0_B1_C0_D1_E1 | 0       | 94         | 11         | 50         | 471        | 626           | <b>0,755</b> | 0,895     | 0,653      |
| 38 | A1_B0_C1_D1_E0 | 1       | 95         | 10         | 52         | 469        | 626           | <b>0,754</b> | 0,905     | 0,646      |
| 39 | A0_B1_C0_D0_E1 | 1       | 93         | 12         | 49         | 472        | 626           | <b>0,753</b> | 0,886     | 0,655      |
| 40 | A1_B1_C1_D1_E1 | 0       | 93         | 12         | 49         | 472        | 626           | <b>0,753</b> | 0,886     | 0,655      |
| 41 | A1_B1_C0_D0_E1 | 0       | 94         | 11         | 51         | 470        | 626           | <b>0,752</b> | 0,895     | 0,648      |
| 42 | A0_B0_C0_D1_E0 | 0       | 95         | 10         | 53         | 468        | 626           | <b>0,751</b> | 0,905     | 0,642      |
| 43 | A1_B0_C0_D1_E1 | 0       | 95         | 10         | 53         | 468        | 626           | <b>0,751</b> | 0,905     | 0,642      |
| 44 | A1_B1_C0_D1_E1 | 0       | 95         | 10         | 53         | 468        | 626           | <b>0,751</b> | 0,905     | 0,642      |
| 45 | A1_B1_C1_D1_E1 | 1       | 93         | 12         | 50         | 471        | 626           | <b>0,750</b> | 0,886     | 0,650      |
| 46 | A0_B0_C0_D0_E1 | 1       | 92         | 13         | 49         | 472        | 626           | <b>0,748</b> | 0,876     | 0,652      |
| 47 | A0_B0_C0_D1_E1 | 0       | 93         | 12         | 51         | 470        | 626           | <b>0,747</b> | 0,886     | 0,646      |
| 48 | A1_B1_C0_D1_E1 | 1       | 97         | 8          | 58         | 463        | 626           | <b>0,746</b> | 0,924     | 0,626      |
| 49 | A0_B1_C0_D1_E1 | 1       | 94         | 11         | 53         | 468        | 626           | <b>0,746</b> | 0,895     | 0,639      |
| 50 | A0_B0_C1_D1_E0 | 0       | 92         | 13         | 50         | 471        | 626           | <b>0,745</b> | 0,876     | 0,648      |
| 51 | A0_B0_C1_D1_E0 | 1       | 92         | 13         | 50         | 471        | 626           | <b>0,745</b> | 0,876     | 0,648      |
| 52 | A1_B0_C1_D1_E0 | 0       | 94         | 11         | 54         | 467        | 626           | <b>0,743</b> | 0,895     | 0,635      |
| 53 | A0_B0_C0_D0_E1 | 0       | 92         | 13         | 51         | 470        | 626           | <b>0,742</b> | 0,876     | 0,643      |
| 54 | A0_B0_C0_D1_E0 | 1       | 96         | 9          | 58         | 463        | 626           | <b>0,741</b> | 0,914     | 0,623      |
| 55 | A0_B1_C0_D1_E0 | 0       | 97         | 8          | 60         | 461        | 626           | <b>0,740</b> | 0,924     | 0,618      |
| 56 | A0_B0_C1_D1_E1 | 1       | 91         | 14         | 50         | 471        | 626           | <b>0,740</b> | 0,867     | 0,645      |
| 57 | A1_B1_C0_D0_E1 | 1       | 93         | 12         | 54         | 467        | 626           | <b>0,738</b> | 0,886     | 0,633      |
| 58 | A1_B1_C1_D1_E0 | 1       | 93         | 12         | 54         | 467        | 626           | <b>0,738</b> | 0,886     | 0,633      |
| 59 | A0_B1_C0_D1_E0 | 1       | 95         | 10         | 58         | 463        | 626           | <b>0,736</b> | 0,905     | 0,621      |
| 60 | A1_B1_C1_D1_E0 | 0       | 93         | 12         | 55         | 466        | 626           | <b>0,735</b> | 0,886     | 0,628      |
| 61 | A0_B1_C1_D1_E0 | 0       | 92         | 13         | 54         | 467        | 626           | <b>0,733</b> | 0,876     | 0,630      |
| 62 | A0_B1_C1_D1_E0 | 1       | 90         | 15         | 51         | 470        | 626           | <b>0,732</b> | 0,857     | 0,638      |
| 63 | A1_B1_C0_D1_E0 | 0       | 94         | 11         | 62         | 459        | 626           | <b>0,720</b> | 0,895     | 0,603      |
| 64 | A1_B1_C0_D1_E0 | 1       | 93         | 12         | 61         | 460        | 626           | <b>0,718</b> | 0,886     | 0,604      |

La nomenclature du prompt indique le modèle de prompt appliqué à chacune des cinq sections (0 ou 1). À chaque annotation est associée une configuration unique (prompt + paramétrage du LLM). Les résultats sont classés par score F1 décroissant. La configuration maximisant le score F1 (n = 1) était utilisée pour application sur l'ensemble de test. spec\_pm : paramétrage du LLM (0 = déterministe, 1 = stochastique) ; confmat\_vp : nombre de vrais positifs ; confmat\_fn : nombre de faux négatifs ; confmat\_fp : nombre de faux positifs ; confmat\_vn : nombre de vrais négatifs ; confmat\_total : nombre total de documents annotés ; metric\_f1 : F-mesure (F1) ; metric\_se : sensibilité ; metric\_vpp : valeur prédictive positive.

**AUTEUR : Nom : ELICES-DIEZ Prénom : Julien**

**Date de Soutenance : 14/10/2025**

**Titre de la Thèse :** Amplification d'un signal de pharmacovigilance à partir d'un entrepôt de données hospitalier : application à la détection des cas d'infections bactériennes associées à la prise d'anti-inflammatoires non stéroïdiens

**Thèse - Médecine - Lille 2025**

**Cadre de classement :** Médecine

**DES + FST ou option :** DES de Santé Publique et Médecine Sociale

**Mots-clés :** pharmacovigilance ; entrepôt de données hospitalier ; anti-inflammatoires non stéroïdiens

**Résumé :**

**Contexte :**

En France, l'incidence des cas infections bactériennes associées à une consommation d'anti-inflammatoires non stéroïdiens (AINS) est vraisemblablement sous-estimée. L'identification automatisée d'effets indésirables médicamenteux à partir de données textuelles issues d'un entrepôt de données hospitalier (EDS) permettrait de faciliter le ciblage de cas éligibles à une analyse d'imputabilité. L'objectif de cette étude était de développer et évaluer une méthode de détection de cas potentiels d'infections bactériennes sous AINS, à partir d'un EDS.

**Matériel et Méthodes :**

L'EDS du CHU de Lille (INCLUDE) était exploité sur la période de janvier et juillet 2018 et 2023 (1<sup>er</sup> au 31 inclus). Les données textuelles en provenance des urgences et d'autres unités d'hospitalisation ont été extraites. Un *large language model* (LLM) était utilisé pour développer une stratégie de détection automatisée de cas éligibles. Le LLM était exécuté à partir d'un *prompt* visant à détecter d'une part un contexte infectieux et d'autre part une consommation récente d'AINS : 6 caractéristiques binaires étaient testées à travers 64 configurations. Les données de 2018 étaient utilisées pour affiner la stratégie ; les données de 2023 étaient utilisées comme échantillon de test. L'annotation automatisée était évaluée par confrontation à une annotation experte avec double lecture indépendante et consensus en cas de désaccord. Le score F1 était estimé.

**Résultats :**

Sur la période étudiée, pour l'échantillon de test, 143 textes médicaux étaient annotés. L'annotation experte retenait 17 (11,9 %) cas éligibles. La proportion d'hommes était de 58,8 % ; la plupart étaient des adultes d'âge médian 42,1 ans (IQR 27,5 – 53,5). La sensibilité était de 88,2 % (15 sur 17 classés éligibles par la référence) et la VPP était de 60,0 % (15 sur 25 classés éligibles par le LLM). Le score F1 était de 71,4 %.

**Conclusion :**

Nous avons proposé une méthode préliminaire de détection automatisée de cas potentiels d'infections bactériennes sous AINS à partir d'un EDS. Les cas identifiés sont éligibles à une analyse d'imputabilité et pourraient aider à la caractérisation du signal concerné. Celle-ci permettrait la réalisation ultérieure d'études pharmaco-épidémiologiques.

**Composition du Jury :**

**Président :** Monsieur le Professeur Philippe AMOUYEL

**Assesseurs :** Monsieur le Professeur Grégoire FICHEUR

Monsieur le Docteur Aghilès HAMROUN

**Directrice :** Madame le Docteur Johana BÉNÉ