

La résistance dans les infections
bactériennes : apport de l'entrepôt de
données de santé de l'AP-HP

Antimicrobial resistant infections:
contribution of the AP-HP Clinical Data
Warehouse

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 570, Santé Publique (EDSP)
Spécialité de doctorat : santé publique - épidémiologie
Graduate School : Santé publique
Réfèrent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche **CESP** (Université Paris-Saclay, UVSQ, Inserm),
sous la direction de **Laurence WATIER**, CR Inserm

Thèse soutenue à Paris, le 17 octobre 2022, par

Salam ABBARA

Composition du Jury

Laurence MEYER PU-PH, Université Paris-Sud (Paris-Saclay)	Présidente
Hugues BERRY DR, INRIA (Lyon)	Rapporteur & Examineur
Elisabeth BOTELHO-NEVERS PU-PH, Université Jean Monnet (Saint-Etienne)	Rapporteuse & Examinatrice
Juliette DJADI-PRAT PH, Université Paris-Cité	Examinatrice
Laurence WATIER CR Inserm, UVSQ (Paris-Saclay)	Directrice de thèse

Titre : La résistance dans les infections bactériennes : apport de l'entrepôt de données de santé de l'AP-HP

Mots clés : Antibiorésistance ; Bactériémie ; Entrepôt de Données de Santé ; Épidémiologie

Résumé : L'objectif de cette thèse était d'évaluer l'apport des données hospitalières pour l'étude des infections bactériennes résistantes aux antibiotiques, à travers l'exemple des bactériémies et de l'Entrepôt de Données de Santé de l'Assistance Publique – Hôpitaux de Paris.

Une base de données clinico-microbiologiques de plus de 30 000 patients hospitalisés avec une bactériémie dans 14 hôpitaux avec une activité de soins aigus entre 2016 et 2019 a été structurée (base BactHub). Un travail exploratoire évalue l'apport d'un programme de Traitement Automatisé du Langage Naturel pour l'identification des sites primaires des bactériémies dans les comptes-rendus, en comparaison avec les codages CIM-10.

Les principales caractéristiques des patients, premiers épisodes et bactéries sont décrites, selon l'origine communautaire ou nosocomiale de la bactériémie. Plus de la moitié des patients étaient de sexe masculin, et d'âge >60 ans. Les taux de mortalité étaient non négligeables, et allaient de 14 à 19% en

intra-hospitalier à 20-26% à J90. Les taux de résistance aux antibiotiques étaient importants : *K. pneumoniae* 3G-R, 21-37% ; *E. coli* 3G-R 13-17% ; SARM 11-14%. La comparaison de nos résultats avec la littérature conforte l'utilisation de la base pour la recherche.

Ensuite, une étude souligne l'impact majeur de l'identification d'une *Klebsiella spp.* 3G-R (OR 4,7), et de manière secondaire d'un *E. coli* 3G-R (OR 2,5), sur le risque de récurrence de bactériémie à 1 an, dans les bactériémies communautaires, après ajustement. A contrario, l'identification d'un SARM n'était pas liée à la récurrence.

Enfin, les points clés et les défis liés à l'étude du lien entre l'exposition individuelle aux antibiotiques et la survenue d'infections bactériennes communautaires résistantes aux antibiotiques sont présentés. Afin d'améliorer et de standardiser les futures études, une proposition est discutée. En conclusion, l'enrichissement de la base avec des données issues du SNDS augmenterait fortement le potentiel de la base pour la recherche.

Title: Antimicrobial resistant infections: contribution of the AP-HP Clinical Data Warehouse

Keywords: Antimicrobial resistance; Bacteremia; Clinical Data Warehouse; Epidemiology

Abstract: The objective of this thesis was to evaluate the contribution of hospital data for the study of antibiotic-resistant bacterial infections, through the example of bacteremia and the clinical data warehouse of the Assistance Publique - Hôpitaux de Paris.

A clinico-microbiological database of more than 30,000 patients hospitalized with bacteremia in 14 hospitals with acute care activity between 2016 and 2019 was structured (the BactHub database). An exploratory work evaluated the contribution of a Natural Language Processing program for the identification of primary sites of bacteremia in medical reports, in comparison with ICD-10 coding.

The main characteristics of the patients, first episodes and bacteria are described, according to the community-onset or nosocomial origin of the bacteremia. More than half of the patients were male and >60 years old. Mortality rates were significant, ranging from 14-19% in the hospital to 20-26% at

D90. Antibiotic resistance rates were significant: 3GC-R *K. pneumoniae*, 21-37%; 3GC-R *E. coli* 13-17%; MRSA 11-14%. The comparison of our results with the literature supports the use of the database for research.

Then, a study underlines the major impact of the identification of a 3GC-R *Klebsiella spp.* (OR 4.7), or a 3GC-R *E. coli* (OR 2.5), on the risk of recurrence of bacteremia at 1 year, in community-onset bacteremia. In contrast, identification of MRSA was not related to recurrence.

Finally, key points and challenges related to the study of the link between individual antibiotic exposure and the occurrence of antibiotic-resistant community-acquired bacterial infections are presented. To improve and standardize future studies, a proposal is discussed. In conclusion, enriching the BactHub database with data from the SNDS would greatly increase its potential for research.



Réseau doctoral
en santé publique



La résistance dans les infections bactériennes : apport de l'entrepôt de données de santé de l'AP-HP

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 570, Santé Publique (EDSP)
Spécialité de doctorat : santé publique - épidémiologie
Graduate School : Santé publique
Référent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche **CESP** (Université Paris-Saclay, UVSQ, Inserm), sous la direction
de **Laurence WATIER**, CR Inserm

Thèse préparée dans le cadre du Réseau doctoral en santé publique animé par l'EHESP

Thèse soutenue à Paris, le 17 octobre 2022, par

Salam ABBARA

Composition du Jury

Laurence MEYER PU-PH, Université Paris-Sud (Paris-Saclay)	Présidente
Hugues BERRY DR, INRIA (Lyon)	Rapporteur & Examineur
Elisabeth BOTELHO-NEVERS PU-PH, Université Jean Monnet (Saint-Etienne)	Rapporteuse & Examinatrice
Juliette DJADI-PRAT PH, Université Paris-Cité	Examinatrice
Laurence WATIER CR Inserm, UVSQ (Paris-Saclay)	Directrice de thèse

PRODUCTIONS SCIENTIFIQUES

Article publié

Abbara S, Guillemot D, Brun-Buisson C, Watier L. From Pathophysiological Hypotheses to Case–Control Study Design: Resistance from Antibiotic Exposure in Community-Onset Infections. *Antibiotics*. 4 febr 2022;11(2):201. DOI: 10.3390/antibiotics11020201

Article soumis

Abbara S, Guillemot D, El Oualydy S, Kos M, Poret C, Breant S, Brun-Buisson C, Watier L. Clinico-microbiological features and mortality of hospitalized bacteremia in the Greater Paris area.

Conférences

Communication orale

Salam Abbara, Didier Guillemot, Christian Brun-Buisson, Laurence Watier. Clinico-microbiological characteristics and mortality of community-onset bacteraemia in French university hospitals, 2016-2019. 32nd European Congress of Clinical Microbiology and Infectious Diseases, Avril 2022, Portugal.

Communications affichées

Salam Abbara, Didier Guillemot, Christian Brun-Buisson, Laurence Watier. Clinico-microbiological characteristics and mortality of hospital-acquired bacteraemia in French university hospitals, 2016-2019. 32nd European Congress of Clinical Microbiology and Infectious Diseases, Avril 2022, Portugal

Salam Abbara, Didier Guillemot, Salma El Oualydy, Maeva Kos, Cécile Poret, Stéphane Breant, Christian Brun-Buisson, Laurence Watier. Caractéristiques clinico-microbiologiques et mortalité des épisodes avec bactériémie nosocomiale dans 14 hôpitaux universitaires en France, 2016-2019. 23^{ème} Journées Nationales d'Infectiologie, Juin 2022, France.

Salam Abbara, Didier Guillemot, Salma El Oualydy, Maeva Kos, Cécile Poret, Stéphane Breant, Christian Brun-Buisson, Laurence Watier. Caractéristiques clinico-microbiologiques et mortalité des épisodes avec bactériémie nosocomiale dans 14 hôpitaux universitaires en France, 2016-2019. 23^{ème} Journées Nationales d'Infectiologie, Juin 2022, France.

Autres valorisations

Salam Abbara. Bacthub : Exposition aux antibiotiques et incidence des bactériémies à bactéries résistantes en France. Bénéfices de l'utilisation des données de santé. Colloque Données de santé et intelligence collective du Health Data Hub, de la Délégation ministérielle du Numérique en Santé, et du Grand Défi « Amélioration des diagnostics médicaux par l'IA ». 9 décembre 2020.

RÉSUMÉ DE LA THÈSE

L'objectif de cette thèse était d'évaluer l'apport des données hospitalières pour l'étude des infections bactériennes résistantes aux antibiotiques, à travers l'exemple des bactériémies et de l'Entrepôt de Données de Santé (EDS) de l'Assistance Publique – Hôpitaux de Paris. Pour répondre à cet objectif, une base de données clinico-microbiologiques de plus de 30 000 patients hospitalisés avec une bactériémie dans 14 hôpitaux universitaires de l'AP-HP avec une activité de soins aigus entre 2016 et 2019 a été structurée (base BactHub). Le processus de construction est présenté, et certains écueils sont soulignés, comme la mauvaise complétude des données d'exposition aux antibiotiques à l'hôpital, ou l'impossibilité de calculer le score de sévérité SOFA de l'infection. Par la suite, un travail exploratoire d'identification des sites primaires des bactériémies dans les comptes-rendus d'hospitalisation (CRH) par traitement automatisé du langage naturel (TAL) est présenté. Ses résultats sont confrontés à une lecture experte des CRH, et aux codages CIM-10 d'infections bactériennes. Le programme de TAL identifiait deux fois plus de sites primaires réels, parmi ceux identifiés, que les codages CIM-10. Il y avait néanmoins une part significative de faux négatifs, les CRH de plus d'un épisode évalué sur 4 ne contenant pas d'information sur l'épisode infectieux. Puis, les principales caractéristiques des patients, des premiers épisodes de bactériémie, et des isolats bactériens dans la base sont décrits, selon le caractère communautaire ou nosocomial de l'épisode. Les patients présentaient un fort taux de mortalité atteignant plus de 20% à J90. Les taux de mortalité associés aux épisodes communautaires étaient plus faibles (~5%) qu'ils soient intra-hospitalier, à J30 ou J90. Les isolats bactériens présentaient de forts taux d'antibiorésistance dans les deux groupes (*K. pneumoniae* C3G-R, 21-37% ; *E. coli* C3G-R 13-17% ; SARM 11-14%), qui variait selon les sites primaires. Une incidence a été estimée, à 20-22 séjours avec bactériémie / 1 000 séjours ; elle augmentait de 2017 à 2019. Ensuite, les facteurs de risque de récurrence de bactériémie à 1 an ont été évalués sur plus de 3 500 épisodes incidents communautaires. Le facteur le plus lié à la récurrence était l'identification d'une *Klebsiella spp.* C3G-R (OR 4,7), puis d'un *E. coli* C3G-R (OR 2,5), alors que l'identification d'un SARM n'était pas liée à la récurrence, après ajustement.

Enfin, une revue des études du lien entre l'exposition individuelle aux antibiotiques et l'occurrence d'infections bactériennes communautaires résistantes aux antibiotiques est présentée, se focalisant sur les infections urinaires et les bactériémies. Les études présentaient une hétérogénéité méthodologique importante, impactant l'interprétation de leurs résultats. Peu d'études se focalisaient sur les infections communautaires vraies, et peu incluaient des témoins non infectés. En chaînant les données de la base BactHub avec des données issues du SNDS, l'exposition aux antibiotiques des patients pourrait être reconstituée, et les bactériémies communautaires vraies pourraient être isolées et étudiées.

« On vous a dit aussi que la vie n'est que ténèbres. Et dans votre lassitude, vous faites écho à ce que disent les éreintés.

Et je vous dis que la vie est ténèbres, si elle est dépourvue d'élan. Et tout élan est aveugle, s'il n'est pas guidé par la connaissance.

Et toute connaissance est vaine, si elle n'est pas appliquée dans un travail.

Et tout travail est futile, s'il n'est pas réalisé avec amour.

Et lorsque vous travaillez avec amour, vous resserrez vos liens avec vous-mêmes, avec autrui et avec Dieu.

Et qu'est-ce que travailler avec amour ?

C'est tisser un vêtement avec des fils tirés de votre cœur, comme si votre bien-aimé devait porter ce vêtement.

C'est construire une maison avec affection, comme si votre bien-aimée devait habiter cette maison.

C'est semer des graines avec tendresse et récolter la moisson avec joie, comme si vos enfants devaient en manger le fruit.

C'est charger toutes choses que vous façonnez d'un souffle de votre propre esprit.

Et savoir que tous les morts bienheureux sont là près de vous et vous regardent. »

Gibran Khalil Gibran, Le Prophète.

TABLE DES MATIÈRES

1	INTRODUCTION.....	11
1.1	Les bactéries et leur hôte : commensalisme, colonisation et infection	11
1.2	Les antibiotiques.....	15
1.3	L'antibiorésistance.....	17
1.4	Sources de données pour l'étude de l'antibiorésistance dans les infections bactériennes	19
2	OBJECTIFS DE LA THÈSE.....	24
3	LES BACTÉRIÉMIES	25
3.1	Intérêt de l'étude des bactériémies.....	25
3.2	Définitions.....	25
3.3	Quelques chiffres.....	27
4	L' ENTREPÔT DES DONNÉES DE SANTÉ DE L'AP-HP.....	29
4.1	Présentation et modalités d'accès	29
4.2	Source et intégration des données	30
4.3	Information des patients.....	33
5	BASE DE DONNÉES BACTHUB	34
5.1	Points techniques	34
5.1.1	Entrepôt de données de santé de l'AP-HP	34
5.1.2	Enrichissement par Traitement Automatisé du Langage : analyse exploratoire	40
5.2	Contrôle qualité	44
5.2.1	Données issues de l'entrepôt de données de santé de l'AP-HP.....	44
5.2.2	Données issues du programme de Traitement Automatisé du Langage	45
6	CARACTÉRISTIQUES, ANTIBIORÉSISTANCE, MORTALITÉ ET INCIDENCE DES BACTÉRIÉMIES... 50	50
6.1	Matériels et méthodes.....	50
6.2	Résultats.....	51
6.3	Discussion.....	59
7	RÉSISTANCE AUX ANTIBIOTIQUES ET RÉCURRENCE DE BACTÉRIÉMIE A 1 AN DANS LES BACTÉRIÉMIES COMMUNAUTAIRES.....	63
7.1	Introduction	63
7.2	Matériels et Méthodes.....	64
7.3	Résultats.....	66
7.4	Discussion.....	71
8	EXPOSITION INDIVIDUELLE AUX ANTIBIOTIQUES ET INFECTIONS BACTÉRIENNES RÉSISTANTES AUX ANTIBIOTIQUES.....	74
8.1	Introduction	74
8.2	Matériels et méthodes.....	75
8.3	Résultats.....	76
8.4	Discussion.....	85
8.5	Apport de la base BactHub.....	92
9	DISCUSSION GÉNÉRALE	94
9.1	Synthèse des principaux résultats.....	94
9.2	Les EDS en épidémiologie	96
10	CONCLUSION GÉNÉRALE	103
11	RÉFÉRENCES.....	104
12	ANNEXES	118

LISTE DES FIGURES

Figure 1.1 Taxonomie bactérienne, exemple pour l'espèce <i>Escherichia coli</i> et la souche ST131	14
Figure 1.2 Principaux modes d'action des antibiotiques et familles associées	16
Figure 1.3 Publicité pour la pénicilline dans le magazine Life, 14 août 1944.....	17
Figure 1.4 Principaux mécanismes de résistance aux antibiotiques.....	19
Figure 4.1 Les données de l'AP-HP, de la source à l'espace projet.....	30
Figure 4.2 État des lieux du déploiement des modules d'Orbis au 09/11/2021	31
Figure 4.3 Cartographie des données de l'EDS de l'AP-HP en date du 16/03/2022	32
Figure 5.1 Exemple d'un séjour, avec plusieurs épisodes de bactériémie.....	38
Figure 5.2 Pré-annotation de concepts dans un document médical de la base BactHub, visualisée sur l'outil brat.....	43
Figure 6.1 Identification des patients avec un épisode de bactériémie et inclusion des données des premiers épisodes, BactHub 2016-2019, groupe communautaire et nosocomial	52
Figure 6.2 Principaux taux de résistance aux antibiotiques des isolats bactériens dans le sang par site d'infection primaire et selon le groupe, base BactHub, 2016-2019.....	58
Figure 7.1 Identification des épisodes incidents communautaires à <i>S. aureus</i> , <i>E. coli</i> ou <i>Klebsiella spp</i> , BactHub 2017-2019, sans décès.....	66
Figure 7.2 Odds ratio ajustés et intervalles de confiance à 95% du lien entre chaque couple bactérie-résistance et la récurrence de bactériémie à 1 an.....	70
Figure 8.1 Diagramme PRISMA de sélection des études cas-témoins évaluant le lien entre l'exposition aux antibiotiques et la survenue d'une IU ou d'une bactériémie communautaire résistante aux antibiotiques	76
Figure 8.2 Hypothèse physiopathologique pour étudier l'exposition aux antibiotiques en tant que facteur de risque d'infection par <i>Escherichia coli</i> ST131 BLSE.....	90

LISTE DES TABLEAUX

Tableau 3.1 Classification des bactériémies.....	27
Tableau 4.1 Intégration des données relatives au circuit du médicament dans l'EDS de l'AP-HP, selon leur source, à la date du 16/03/2022 [95]	32
Tableau 5.1 Valeurs minimales et maximales des taux de remplissage (%) des variables physiologiques et biologiques nécessaires pour le calcul du score de SOFA sur l'ensemble des séjours avec bactériémie communautaire et sur ceux avec un passage en réanimation - BactHub 2016-2019.....	39
Tableau 5.2 Hôpitaux avec une activité de soins aigus à l'AP-HP	45
Tableau 5.3 Identification des bactériémies et des sites primaires dans les CRH d'épisodes de bactériémie par le programme de TAL ou les codages CIM-10, en comparaison avec une lecture experte	47
Tableau 5.4 Répartition des sites primaires pour 216 épisodes de bactériémie, selon la technique d'identification des sites primaires, en pourcentage (%).....	49
Tableau 6.1 Caractéristiques des patients selon le groupe, base BactHub, 2016 – 2019	53
Tableau 6.2 Caractéristiques des épisodes de bactériémie selon le groupe, base BactHub, 2016 – 2019.	54
Tableau 6.3 Distribution des isolats bactériens selon le groupe, base BactHub, 2016 – 2019.....	55
Tableau 6.4 Principaux taux de résistance aux antibiotiques (en pourcentage) des isolats bactériens selon le groupe, base BactHub, 2016 – 2019.....	57
Tableau 7.1 Caractéristiques des patients et des épisodes incidents communautaire à S. aureus, E. coli ou Klebsiella spp, avec et sans récurrence, base BactHub, 2017-2019	68
Tableau 7.2 Analyses univariées et multivariées des facteurs de risque de récurrence à 1 an suite à un épisode incident communautaire monomicrobien à S. aureus, E. coli ou Klebsiella spp	70
Tableau 8.1 Classification, fenêtre d'exposition, et source des données sur l'exposition aux antibiotiques dans les études cas-témoins évaluant les facteurs de risque d'antibiorésistance dans les bactériémies et les infections urinaires communautaires.....	78
Tableau 8.2 Définitions des infections communautaires dans les études cas-témoins évaluant les facteurs de risque d'antibiorésistance dans les bactériémies et les infections urinaires	79
Tableau 8.3 Principaux résultats des études cas-témoins évaluant les facteurs de risque d'antibiorésistance dans les bactériémies communautaires à entérobactéries BLSE, selon le groupe témoin	82
Tableau 8.4 Principaux résultats des études cas-témoins évaluant les facteurs de risque d'antibiorésistance dans les infections urinaires communautaires à entérobactéries BLSE, selon le groupe témoin.....	83
Tableau 8.5 Points clés pour concevoir des études cas-témoins évaluant l'exposition aux antibiotiques comme facteur de risque d'infections communautaires résistantes aux antibiotiques.....	88
Tableau 8.6 Sélection des groupes cas et témoins en fonction de la question spécifique	89

LISTE DES ABRÉVIATIONS

AP-HP : Assistance Publique – Hôpitaux de Paris
BHRe : bactérie hautement résistante aux antibiotiques émergente
BLSE : β -lactamases à spectre étendu
BMR : bactérie multi-résistante aux antibiotiques
C3G : céphalosporines de troisième génération
C3G-R : résistant(e) aux C3G
CCAM : classification commune des actes médicaux
CHU : centre hospitalier universitaire
CPias : centres d'appui pour la prévention des infections associées aux soins
CRH : compte-rendu d'hospitalisation
CSE : comité scientifique et éthique
DAS : diagnostic associé significatif
DDJ : Doses Définies Journalières
DPI : dossier patient informatisé
DP : diagnostic principal
DR : diagnostic relié
E-BLSE : entérobactéries productrices de BLSE
ECDC : centre européen de prévention et de contrôle des maladies
EDS : entrepôt de données de santé
ERV : *Enterococcus faecium* résistant à la vancomycine
ETL : « extraction, transformation, loading »
HAD : hospitalisations à domicile
HDH : Health Data Hub
IAS : infection associée aux soins
IC95 : intervalle de confiance à 95%
MCO : médecine, chirurgie et obstétrique
PMSI : Programme de Médicalisation des Systèmes d'Information
PACS : picture archive and communication system
PRIMO : surveillance et prévention de la résistance aux antibiotiques et des infections associées aux soins en soins de ville et en secteur médico-social
Q1, Q3 : premier quartile, troisième quartile
SARM : *S. aureus* résistant à la méticilline
SI : système d'information
S, I, R : sensible, intermédiaire, résistant
SNDS : Système National des Données de Santé
SNIIRAM : Système National d'Information Inter Régimes de l'Assurance Maladie
SPARES : surveillance et prévention de la résistance bactérienne aux antibiotiques en établissement de santé
SpF : Santé publique France
SSR : soins de suite et réadaptation
TAL : traitement automatisé du langage naturel
VP, FP, VN, FN : vrai positif, faux positif, vrai négatif, faux négatif

1 INTRODUCTION

« Si la chaleur innée est faible, les facultés naturelles sont harcelées dans la régulation des humeurs. Car l'instrument intermédiaire entre les facultés naturelles et les humeurs est affaibli. La stagnation s'installe et la chaleur étrangère ne trouve plus les humeurs opposées à son action. Elle les surmonte. Elle les utilise à sa manière et leur imprime un mouvement étranger ; il en résulte ce qu'on appelle la « putréfaction » ».

Avicenne, *Canon de la Médecine*, 1020.

« La terre est un grand estomac, dans lequel tout est dissout, digéré, et transformé, et chaque être tire ses nutriments de la terre ; et chaque être vivant est un estomac qui sert de tombeau à d'autres formes, et d'où surgissent de nouvelles formes. »

Paracelse, *Paramirum*, 1575.

Si l'épidémiologie est une discipline qui étudie les problèmes de santé et leur distribution dans la population [1], elle se nourrit d'une compréhension à l'échelle micro du fonctionnement et des dynamiques internes à chaque être humain et d'autres organismes, comme – dans le cadre de cette thèse – les bactéries, ainsi que leurs interactions.

1.1 Les bactéries et leur hôte : commensalisme, colonisation et infection

Les bactéries sont des micro-organismes unicellulaires, qui à la différence des cellules humaines ne possèdent pas de noyau (procaryotes, vs eucaryotes). Elles se reproduisent de manière asexuée, par fission binaire, c'est-à-dire qu'elles grandissent et se divisent en deux après une simple duplication de leur ADN et de leurs autres constituants. Ces micro-organismes micrométriques colonisaient la Terre depuis bien plus longtemps que nous, et ont su s'adapter et évoluer dans des environnements extrêmement difficiles auxquels peu d'eucaryotes auraient survécu. Apparues au plus tard il y a 3.5 milliards d'années [2], elles ont joué un rôle majeur dans l'évolution de la Terre, d'un vaste océan avec une atmosphère sans oxygène, vers un espace habitable pour l'être humain [3–6]. Si l'on se focalisera dans cette thèse sur les bactéries dans leur lien avec les êtres humains et les infections, il est bon de rappeler qu'elles sont partout, dans tous les écosystèmes, en entretenant des relations très diverses aux plantes, aux animaux, à l'être humain. A titre d'exemple, « il y a 100 millions de fois plus de bactéries dans les océans que d'étoiles connues à ce jour dans l'univers » [7].

Pathogénicité

Depuis les premières formes de vie humaine, l'être humain a appris à vivre avec les bactéries, et interagit avec elles de manière complexe. Une notion clé pour comprendre les relations entre l'être humain et les bactéries est le pouvoir pathogène des bactéries, c'est-à-dire leur capacité à provoquer des troubles chez leur hôte. Ces troubles peuvent résulter de l'action directe des bactéries et/ou de la réponse immunitaire et inflammatoire de l'hôte¹ [8]. Le pouvoir pathogène des bactéries est intimement lié à : (1) leur pouvoir adhésif (étape obligatoire d'adhésion au niveau de la peau ou des muqueuses) ; (2) leur pouvoir invasif (capacité à se multiplier dans les tissus et à y développer un foyer infectieux) ; (3) leur pouvoir toxigène (capacité à produire des toxines dont les effets vont s'exercer à distance de la porte d'entrée) ; et (4) leur mode de multiplication (intra ou extracellulaire, entraînant des relations différentes avec le système immunitaire). Ainsi, certaines bactéries peuvent être des pathogènes stricts, c'est-à-dire qu'elles ont le pouvoir d'infecter un hôte qui peut produire une réponse immunitaire normale, comme *Salmonella typhi* (qui donne la fièvre typhoïde), *Neisseria meningitidis* (la méningite à méningocoque), etc. D'autres bactéries, comme *Staphylococcus aureus* ou *Pseudomonas aeruginosa*, sont dites pathogènes opportunistes, c'est-à-dire qu'elles pourront infecter un hôte si ses défenses sont altérées. Selon les bactéries, l'altération des défenses de l'hôte peut par exemple prendre la forme d'une effraction cutanée (et donc le franchissement de la barrière mécanique et immunitaire qu'offre la peau) ou d'une maladie chronique qui affaiblit globalement le système immunitaire du patient (comme les cancers ou le diabète, etc.).

Commensalisme

Si l'on rencontre des bactéries dans tous les environnements avec lesquels on interagit (y compris dans l'air), nous en abritons également beaucoup. Ainsi, dans chaque corps humain cohabitent près de 30 000 milliards de cellules humaines (probablement moins chez les femmes, principalement du fait d'un nombre de globules rouges plus bas), et plus de 38 000 milliards de bactéries (le microbiote), logeant principalement dans le tube digestif, dans la peau et dans les cavités buccales, nasopharyngées, et vaginales [9]. La plupart des bactéries que le corps humain abrite sont acquises à la période néonatale et appelées « bactéries commensales ». L'équilibre entre le corps humain et les bactéries commensales est complexe et se base sur une symbiose, c'est-à-dire une relation de donnant à donnant. D'un côté, nous fournissons aux bactéries commensales un habitat ; de l'autre, elles nous fournissent des nutriments,

¹ Réponse immunitaire : mécanismes permettant à l'hôte de se défendre contre un signal de danger (antigène). La réaction inflammatoire est une composante de la réponse immunitaire ; elle peut être aiguë, subaiguë, chronique ; elle peut être locale (limitée aux tissus atteints, pouvant se manifester par une rougeur, une chaleur, un œdème, et une douleur du site de l'inflammation) ou générale (pouvant donner de la fièvre, etc.).

aident notre métabolisme, modulent notre système immunitaire. Par ailleurs, elles jouent un rôle barrière face à des micro-organismes pathogènes extérieurs à l'organisme (d'où le terme fréquemment employé de « flores barrières »), en colonisant leurs principales portes d'entrée et en produisant des molécules bactéricides [10,11]. Ainsi, le corps ne tente pas d'éliminer ces bactéries à l'état d'équilibre, et leur présence dans l'organisme n'induit aucun symptôme particulier. Cependant, plusieurs éléments peuvent rompre cet équilibre, pouvant participer au développement de maladies inflammatoires, tumorales ou infectieuses. A ce titre, l'étude des liens entre le microbiote intestinal et diverses maladies chroniques ou inflammatoires est un domaine de recherche en plein essor. Par ailleurs, certaines bactéries commensales peuvent devenir pathogènes et infecter leur hôte, en cas d'altération de la muqueuse, de déséquilibre de la flore microbienne, et/ou d'altération des défenses immunitaires de l'hôte, etc. [12]. A titre d'exemple, l'altération de la muqueuse peut avoir comme conséquence le passage d'une bactérie commensale du tube digestif dans le sang, ce que l'on appelle une translocation bactérienne.

Colonisation bactérienne

En plus des bactéries commensales, le corps humain peut être colonisé par des bactéries, acquises n'importe quand durant la vie d'un individu [12]. Ainsi, la colonisation (ou portage) désigne la présence d'une bactérie (espèce, ou souche) au sein de flores bactériennes du corps humain où elle ne devrait pas se trouver (par exemple, la colonisation des voies aériennes supérieures par *Streptococcus pneumoniae*). A la différence du commensalisme, la colonisation déclenche une réaction immunitaire asymptomatique, et la colonisation persiste habituellement jusqu'à élimination par le système immunitaire ou par une antibiothérapie. Dans certains cas, les bactéries colonisatrices peuvent être pathogènes et provoquer une infection chez leur hôte.

Taxonomie bactérienne

Les bactéries étant très nombreuses, il est important d'être rigoureux dans la taxonomie² utilisée pour les caractériser. Les bactéries sont classées de manière hiérarchique en plusieurs niveaux, sur la base de données phylogénétiques³ (**Figure 1.1**). L'unité de base de cette classification est l'espèce : une espèce regroupe des organismes possédant de nombreuses propriétés stables communes. Au sein d'une même espèce, plusieurs sous-divisions, que l'on appelle souches, peuvent exister.

Tout au long de ce manuscrit, j'utiliserai ces quatre niveaux de classification (par ordre décroissant) : la **famille**, le **genre**, **l'espèce**, et la **souche**. A titre d'exemple, je pourrai parler de la **famille** des entérobactéries (qui regroupe des bactéries souvent rencontrées dans les tubes digestifs des humains et

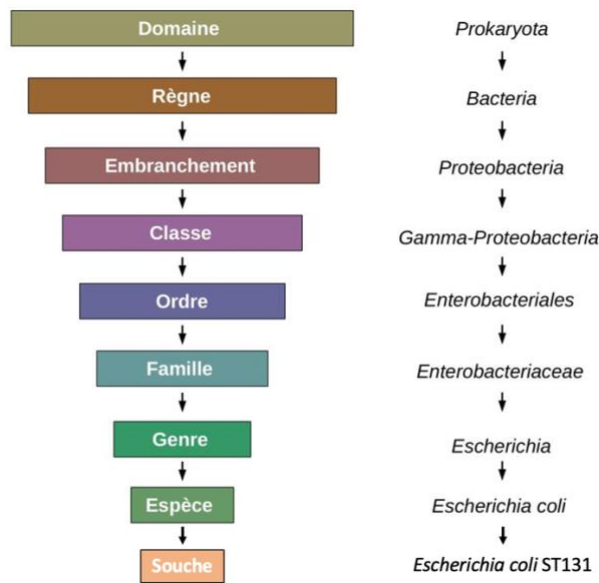


² Taxonomie : science de la classification des organismes vivants

³ Phylogénie : science de l'évolution génétique dans le temps d'une bactérie parmi toutes les autres. Elle permet de regrouper les organismes selon leurs liens de parenté.

des animaux). Dans cette famille, l'on peut isoler entre autres le **genre** *Escherichia*, dont l'**espèce** principale est *Escherichia coli*, dont certaines **souches** sont commensales du tube digestif humain. Parmi les **souches** non commensales d'*E. coli*, certaines sont pathogènes pour l'être humain, pouvant causer des manifestations intestinales (par exemple, *E. coli* entérotoxigène, une des principales causes de diarrhée du voyageur) et/ou extra-intestinales (par exemple, *E. coli* uropathogène) [13–15]. Au sein de la **famille** des entérobactéries, l'on peut aussi retrouver, par exemple, les **genres** *Enterobacter* ou *Klebsiella*, dont l'espèce la plus souvent isolée est *Klebsiella pneumoniae*.

Figure 1.1 Taxonomie bactérienne, exemple pour l'espèce *Escherichia coli* et la souche ST131 [16]



Infection bactérienne

L'être humain peut être infecté par une bactérie de manière endogène, à partir de ses flores bactériennes (commensales ou colonisatrices), ou de manière exogène, en interagissant avec son environnement (air, surfaces, eau, sol, alimentation, etc.) ou d'autres êtres vivants (plantes, insectes, mammifères, etc.), notamment une personne porteuse d'une bactérie pathogène. Par la suite, je définirai l'infection comme la présence d'une maladie induite par une bactérie, se traduisant par divers symptômes cliniques (douleur, rougeur, pus, etc.) auxquels peut s'associer une fièvre.

Schématiquement, la survenue ou non d'une infection dépend des dynamiques intra-hôte opposant d'un côté l'action des défenses immunitaires spécifiques et non spécifiques, de l'autre les facteurs de virulence que les bactéries utilisent pour échapper à ces défenses, séquestrer le fer et d'autres nutriments essentiels à leur réplication pour provoquer des maladies [17]. L'ensemble de ces processus ont lieu de manière continue dans nos organismes, et aboutissent, la plupart du temps, à une situation de non-infection [17].

En cas d'infection, notons que la réaction immunitaire et inflammatoire est une arme à double tranchant, qui peut d'un côté nous protéger et guérir l'infection (parfois sans intervention médicale), mais aussi détruire de manière plus ou moins massive les cellules de l'hôte et parfois mener au décès [17].

1.2 Les antibiotiques

Présentation et classification

Les infections bactériennes sont associées à une morbidité et une mortalité importante. Le spectre de ces infections⁴ et de leur gravité est très large, allant de la simple folliculite bactérienne à la gravissime méningococcémie. Comme indiqué précédemment, la gravité des infections dépend également de la situation des défenses immunitaires de l'hôte. Si certaines infections peuvent guérir sans intervention médicale, comme les cystites, d'autres nécessitent une prise en charge rapide et adaptée. A ce titre, les antibiotiques sont le principal traitement des infections bactériennes. Découverts en 1928, et utilisés à partir des années 40, les antibiotiques sont des substances d'origine naturelle ou synthétique capables d'inhiber (activité bactériostatique) ou de détruire (activité bactéricide) certaines espèces bactériennes, en interférant avec leur cycle répliatif. Ils sont principalement classés selon leur **structure chimique**, leur **mode d'action**, leur **spectre d'action**, et leur caractéristiques **pharmacocinétiques** et **pharmacodynamiques**.

La **structure chimique** des antibiotiques est très variée, et est à la base de leur classification en familles. Par exemple, les macrolides possèdent un noyau lactone central ; les β -lactamines possèdent un cycle β -lactame, etc.

Les principaux **modes d'actions** des antibiotiques et les principales familles d'antibiotiques associées sont représentés dans la **Figure 1.2**.

Le **spectre d'action** d'un antibiotique regroupe l'ensemble des bactéries sur lesquelles il est habituellement actif, permettant de guider son utilisation en pratique clinique. Il peut être plus ou moins étroit, selon le potentiel d'activité plus ou moins large d'un antibiotique sur les différentes espèces bactériennes.

Les principales **caractéristiques pharmacocinétiques** d'un antibiotique sont son mode d'absorption et sa biodisponibilité, son volume de distribution, sa demi-vie dans le sang, et son mode d'élimination.

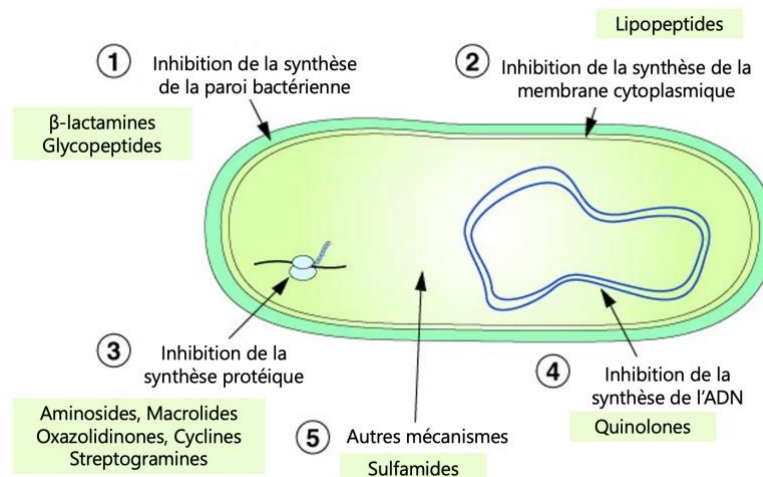
Enfin, les principales **caractéristiques pharmacodynamiques** d'un antibiotique sont : (1) sa bactéricidie (activité bactériostatique ou bactéricide) ; (2) son activité, temps-dépendante (c'est-à-dire fonction de la durée d'exposition des bactéries à l'antibiotique) ou concentration-dépendante (fonction de la concentration d'exposition des bactéries à l'antibiotique) ; (3) son effet post-antibiotique (effet persistant

⁴ Ensemble des infections existantes

sur les bactéries après arrêt de l'exposition à l'antibiotique) ; (4) sa fenêtre de sélection de mutants résistants (concentrations d'antibiotiques pouvant entraîner la sélection de mutants bactériens résistants).

A titre d'exemple, les **β -lactamines** sont la plus importante famille d'antibiotiques à la fois en termes de nombre de molécules et de consommation (57% des consommations d'antibiotiques en France en 2019, en ville et à l'hôpital [18]). Elles partagent une même structure chimique, et agissent en se liant aux protéines de liaison aux pénicillines (PLP), qui sont les enzymes synthétisant le peptidoglycane, principal constituant de la paroi bactérienne. Les β -lactamines ont un spectre d'action large (variable selon les molécules), sont habituellement bactéricides, et présentent une activité temps-dépendante avec un effet post-antibiotique faible ou nul (sauf les carbapénèmes). Au sein de cette famille, on trouve par exemple l'amoxicilline, l'oxacilline, les céphalosporines de troisième génération (C3G), les carbapénèmes, etc [19].

Figure 1.2 Principaux modes d'action des antibiotiques et familles associées [20]



Usage en France

Depuis la découverte des antibiotiques au milieu du XX^{ème} siècle (**Figure 1.3**), leur développement a été considérable et leur utilisation croissante [21].

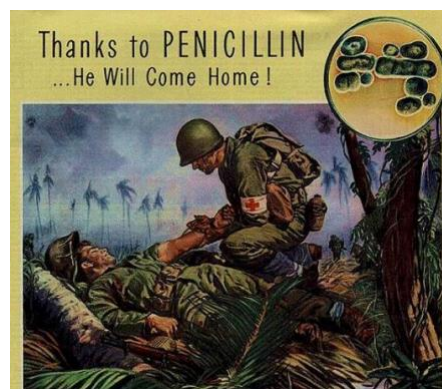
La France fait partie des pays les plus consommateurs d'antibiotiques. En 2019, elle était le 3^{ème} plus gros consommateur d'antibiotiques parmi les pays membres⁵ du Centre européen de prévention et de contrôle des maladies (ECDC), avec une consommation globale (en ville et à l'hôpital) de 25 Doses

⁵ Pays de l'Union Européenne (Allemagne, Autriche, Belgique, Bulgarie, Chypre, Croatie, Danemark, Espagne, Estonie, Finlande, France, Grèce, Hongrie, Irlande, Italie, Lettonie, Lituanie, Luxembourg, Malte, Pays-Bas, Pologne, Portugal, République tchèque, Roumanie, Slovaquie, Slovénie, Suède) et de l'Espace Économique Européen (Islande, Liechtenstein, Norvège).

Définies Journalières⁶ (DDJ) / 1000 habitants par jour, pour une consommation moyenne dans les pays membres de l'ECDC de 19,4 DDJ / 1000 habitants par jour [22].

En France, la plupart des antibiotiques (93% en 2019) sont dispensés en médecine de ville, et une minorité (7%) sont administrés dans les établissements de santé. Quantitativement, en 2019, la consommation des antibiotiques se chiffrait à 22,2 DDJ / 1 000 habitants et par jour en ville, et 286 DDJ / 1 000 journées d'hospitalisation dans les établissements de santé [23–25]. Une part importante des patients pris en charge dans les établissements de santé reçoivent des antibiotiques, cette part pouvant atteindre 30% des patients dans une étude européenne [26].

Figure 1.3 Publicité pour la pénicilline dans le magazine Life, 14 août 1944



1.3 L'antibiorésistance

Un problème de santé publique mondial

La résistance des bactéries aux antibiotiques est un phénomène mondial, en développement continu, qui représente un enjeu majeur de santé publique car peut conduire à des échecs de traitement et une morbi-mortalité accrue [27,28]. A l'hôpital, en France, le nombre de nouveaux cas d'infection à bactérie résistante a été estimé à 139 105 (intervalle de confiance à 95% (IC95), 127 920 – 150 289) en 2016, soit 12% de toutes les infections bactériennes ayant nécessité une hospitalisation, et entraînant un surcoût de 290 millions d'euros [29,30]. Dans les pays membres de l'ECDC, Cassini et al. ont estimé à 671 689 (IC95, 583 148 – 763 966) le nombre d'infections à bactérie résistante en 2015, à partir des données de l'EARS-Net⁷. Par ailleurs, ces infections auraient causé 874 541 (768 837 – 989 068) années de vie

⁶ La DDJ est fixée par l'Organisation Mondiale de la Santé. Pour chaque molécule antibiotique, elle correspond à sa dose moyenne quotidienne pour un adulte de soixante-dix kilos, dans son indication principale. Elle ne reflète pas nécessairement la posologie recommandée par l'autorisation de mise sur le marché (AMM) en France, ni la posologie effective prescrite aux patients (qui tient compte de l'infection exacte et de sa sévérité, et d'autres critères comme le poids, l'âge, les comorbidités comme l'insuffisance rénale, etc).

⁷ Réseau européen de surveillance de la résistance aux antimicrobiens.

corrigées de l'incapacité (AVCI)⁸, avec un nombre de décès attribuables en 2015 de 33 110 (IC95 28 480 – 38 430) [31]. Dans le Monde, en 2019, Murray et al. ont estimé le nombre de décès associés à l'antibiorésistance à 4,95 millions (IC95 3,62 – 6,57), et le nombre de décès attribuables à l'antibiorésistance à 1,27 millions (0,911 – 1,71) [28].

Types et mécanismes de résistance

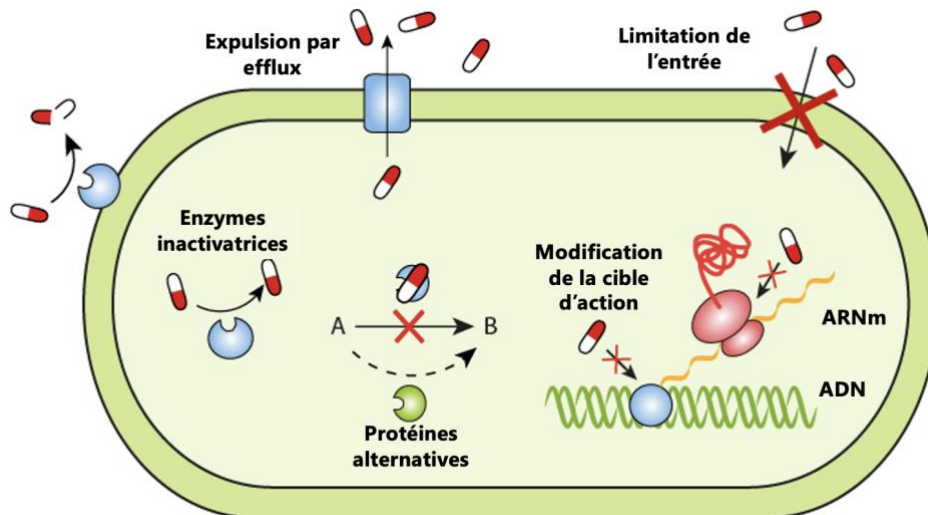
Il existe deux types de résistance aux antibiotiques : (1) les **résistances naturelles**, de certaines espèces à certains antibiotiques, de support chromosomique (par exemple : résistance de *Klebsiella pneumoniae* à l'amoxicilline), (2) les **résistances acquises**, développées par les bactéries pour échapper aux antibiotiques. Ces résistances sont de support majoritairement plasmidique (les plasmides sont des fragments d'ADN circulaire présents dans la cellule bactérienne et indépendants du génome bactérien, transmissibles entre les bactéries) [32], et minoritairement chromosomique (par mutation ou incorporation d'un gène sur le chromosome bactérien, en général non transférable à d'autres bactéries). Les résistances acquises peuvent conduire à (**Figure 1.4**) : (1) inactiver l'antibiotique (par sécrétion d'une enzyme, comme les β -lactamases à spectre étendu (BLSE), qui dégradent la plupart des β -lactamines) ; (2) modifier sa cible d'action (par exemple, *Enterococcus faecium* devient résistant à la vancomycine (ERV) en produisant des précurseurs de la paroi modifiés) ; (3) produire des protéines alternatives à la cible d'action (comme c'est le cas avec le *S. aureus* résistant à la méticilline (SARM) : l'acquisition du gène *mecA* lui permet de fabriquer une version alternative de la cible des β -lactamines, la protéine de liaison à la pénicilline, avec laquelle les β -lactamines ont une faible affinité) ; (4) limiter son entrée dans la bactérie (en diminuant la perméabilité membranaire à l'antibiotique) ; (5) expulser l'antibiotique hors de la bactérie par efflux (comme l'hyperexpression de MexAB-oprM chez *Pseudomonas aeruginosa*, qui conduit à l'expulsion des β -lactamines en dehors de la bactérie) [19].

Les résistances naturelles et acquises ne sont pas mutuellement exclusives et les résistances acquises peuvent s'accumuler chez une bactérie donnée, limitant les options de traitement. Ainsi, les bactéries sont dites multi-résistantes aux antibiotiques (BMR) lorsque, « du fait de l'accumulation des résistances naturelles et acquises, elles ne sont plus sensibles qu'à un petit nombre d'antibiotiques habituellement actifs en thérapeutique » [33]. Les BMR qui font l'objet d'une surveillance actuellement sont les entérobactéries productrices de BLSE (E-BLSE) et le SARM [34]. La dénomination de bactéries hautement résistantes aux antibiotiques émergentes (BHRe) est réservée aux entérobactéries productrices de carbapénémase (EPC) et aux ERV. Cette distinction est opérée pour souligner la capacité de ces bactéries

⁸ L'indicateur AVCI est utilisé pour mesurer le fardeau d'une maladie, et prenant en compte à la fois la mortalité et la perte de santé fonctionnelle. Il correspond à la somme des années de vie perdues et des années vécues avec de l'incapacité.

à s'implanter dans le microbiote intestinal, et le mécanisme plasmidique (donc transférable) de leur résistance [35].

Figure 1.4 Principaux mécanismes de résistance aux antibiotiques [36]



Mécanismes de développement

Le développement de résistances aux antibiotiques est un processus multifactoriel, qui implique : (1) des mutations naturelles, et des transmissions horizontales de matériel génétique entre les bactéries ; (2) l'exposition aux antibiotiques ; (3) la transmission de bactéries résistantes entre humains, animaux, et leur dissémination dans l'environnement [37]. Particulièrement, le lien entre l'exposition aux antibiotiques et le développement de la résistance aux antibiotiques (qu'il s'agisse de colonisation ou d'infection) a été largement documenté in vitro et in vivo [38,39]. Ainsi, le bon usage des antibiotiques est au cœur de tous les programmes de lutte contre l'antibiorésistance [40–42].

1.4 Sources de données pour l'étude de l'antibiorésistance dans les infections bactériennes



Pour étudier la résistance bactérienne dans les infections bactériennes, des informations détaillées cliniques et microbiologiques sur l'infection, incluant des données sur la résistance des isolats bactériens, sont indispensables. Outre les données issues d'études interventionnelles et, par-delà, collectées dans un cadre expérimental, les données recueillies en vie réelle, reflétant la pratique courante, offrent une bonne alternative [43]. Ces données peuvent être issues d'études observationnelles prospectives (nécessitant une interaction avec les patients ; rentrant dans le cadre de la recherche impliquant la

personne humaine dans la loi Jardé) [44–47] ou rétrospectives (sans interaction avec les patients ; hors loi Jardé) [48–50]. Ces dernières peuvent provenir de différentes sources : dossiers informatisés de patients, remboursement de soins, recueil dans le cadre de procédures de pharmacovigilance ou pour constituer des registres, etc. [43]. Pour étudier les infections bactériennes résistantes aux antibiotiques dans le cadre d'une recherche n'impliquant pas la personne humaine, trois sources de données de vie réelle sont intéressantes.

Les réseaux de surveillance en France

Jusqu'en 2018, la surveillance de la résistance aux antibiotiques en établissements de santé reposait sur le réseau BMR-Raisin, qui ciblait le SARM et les E-BLSE isolés de prélèvements à visée diagnostique. La surveillance de la résistance aux antibiotiques en ville reposait, elle, sur des données régionales [34]. En 2018, cinq missions nationales ont été définies par un comité d'expert « Missions nationales Infections associées aux soins », et ont été déléguées pour la période 2018-2023 à 6 centres d'appui pour la prévention des infections associées aux soins (CPias).

Parmi ces missions, la **mission SPARES** « surveillance et prévention de la résistance bactérienne aux antibiotiques en **établissement de santé** » a été confiée par Santé publique France (SpF) au CPias Grand-Est, associé au CPias Nouvelle Aquitaine [34]. Elle a pour objet la surveillance de la consommation des antibiotiques et des résistances bactériennes en établissements de santé ; tous les établissements de santé volontaires peuvent participer [51–53]. En 2020, 1066 établissements de santé participaient à la surveillance de l'antibiorésistance, représentant environ un tiers des établissements de santé en France, la moitié des lits d'hospitalisation, et la moitié des journées d'hospitalisation en France [53]. La surveillance concerne toute bactérie isolée de prélèvements à visée diagnostique et ayant fait l'objet d'un antibiogramme. Ainsi, les prélèvements à visée de recherche de colonisation, portage, ou à visée de dépistage sont exclus. La mission **PRIMO** « surveillance et prévention de la résistance aux antibiotiques et des infections associées aux soins en **soins de ville et en secteur médico-social** » a été confiée par SpF au CPias Pays de la Loire, associé au CPias Grand-Est [54–56]. Cette surveillance repose sur le volontariat des laboratoires de biologie médicale en ville ; le nombre de participants est passé de 742 à 1311 entre 2018 et 2020, conduisant en 2020 à un taux de couverture nationale d'environ 30% [54,56]. Le taux de participation régionale des laboratoires à la mission PRIMO varie selon les régions, entre 0% et 92%. La surveillance concerne les prélèvements à visée diagnostique positifs à *Escherichia coli*, *Klebsiella pneumoniae*, *Staphylococcus aureus*, *Klebsiella oxytoca*, *Enterobacter aerogenes*, et *Enterobacter cloacae* [57]. A la différence des données recueillies dans la mission SPARES, la surveillance se limite à quelques sites de prélèvements, selon la bactérie (par exemple, les urines pour *E. coli* et *K. pneumoniae*).

Globalement, si ces réseaux communiquent une information riche et détaillée pour la surveillance de l'antibiorésistance, leur fiabilité et la représentativité des données recueillies est à vérifier. La participation des établissements de santé et des laboratoires de ville étant basée sur le volontariat, cela peut introduire un biais de sélection si la représentativité des participants n'est pas étudiée. Par ailleurs, l'absence d'étape de validation des informations remontées grâce aux outils de ces réseaux compromet la fiabilité des données. En effet, un établissement de santé ou laboratoire participant pourrait avoir une remontée incomplète, altérée, ou hétérogène dans le temps. Dans la mission PRIMO, les sites de prélèvements analysés étant limités, les résultats ne sont pas transposables aux bactéries issues d'autres sites de prélèvements. Enfin, notons que ces réseaux s'appuient sur des données microbiologiques, et ne donnent aucune information clinique permettant de caractériser la population d'où sont issus les prélèvements.

Le Système National des Données de Santé (SNDS)



Le SNDS est un entrepôt de données médico-administratives pseudonymisées couvrant l'ensemble de la population française et contenant l'ensemble des soins de ville et de l'hôpital présentés au remboursement, ainsi que les causes médicales de décès [58]. Il chaîne trois bases de données : le Système National d'Information Inter Régimes de l'Assurance Maladie (SNIIRAM), les données des hôpitaux et autres établissements de santé (Programme de Médicalisation des Systèmes d'Information - PMSI), et les données statistiques relatives aux causes de décès (BCMD) [59]. Les informations concernant le bénéficiaire dans le SNIIRAM comportent l'âge, le sexe, le lieu de résidence (département et commune), l'indice de défavorisation de la commune d'origine, le régime d'affiliation, la notion éventuelle de bénéfice de la couverture maladie universelle, d'affection de longue durée (ALD) ou de maladie professionnelle, etc. Le SNIIRAM inclut également les informations concernant les soins et actes délivrés du secteur libéral (médicaments, examens complémentaires, consultations médicales, actes techniques, dispositifs médicaux, soins paramédicaux selon le type de professionnel de santé), incluant leur date, et les montants remboursés par l'assurance maladie et payés par le patient. Le PMSI couvre l'ensemble des hospitalisations en médecine, chirurgie et obstétrique (MCO), en soins de suite et réadaptation (SSR), en psychiatrie (RIM-P) et les hospitalisations à domicile (HAD). Le recueil systématique inclut des informations médicales (diagnostics, actes, etc.), des informations de type administratif (identification de l'établissement, durée de séjour, mode d'entrée et de sortie, décès éventuel) et sur le patient (sexe, âge, code géographique de résidence basé sur le code postal de résidence). Les établissements délivrent également des informations sur le nombre de consultations et d'actes externes réalisés et sur la consommation des médicaments et dispositifs facturés « en sus » des forfaits hospitaliers [60]. Ainsi, le SNDS fournit l'exhaustivité des patients hospitalisés avec un code CIM-10 d'infection bactérienne, leurs caractéristiques sociodémographiques, médicales et socioéconomiques et toutes les consommations de soins en ville et à l'hôpital. Cependant, les résultats biologiques (notamment

microbiologiques) ne sont pas disponibles. A partir des données de codage CIM-10 du PMSI, Opatowski et al. ont montré qu'en 2016, l'information sur la bactérie était disponible pour 31% des séjours hospitaliers avec une infection bactérienne aiguë [29]. Ce pourcentage variait selon le site infectieux, par ailleurs il n'était pas possible de dater précisément les prélèvements et donc de déterminer si les infections étaient communautaires ou associées aux soins. De plus, les traitements reçus dans les établissements de santé (à l'exception de ceux de la liste « en sus », cf. ci-dessus) et les données physiologiques, d'examen physique, radiologiques ne sont pas recueillis dans le SNDS, empêchant d'établir la sévérité des infections et de diagnostiquer certaines infections.

Les entrepôts de données de santé

Depuis les années 1960, le développement de l'informatique dans le domaine de la santé a mené au développement et à l'utilisation progressive du dossier patient informatisé (DPI) et de systèmes informatiques conçus pour recueillir, stocker et manipuler les informations médicales et administratives des patients, appelés systèmes d'information (SI) [61–66]. Ces avancées permettent un meilleur partage et traitement de l'information médicale, dans l'objectif d'optimiser la prise en charge du patient. En parallèle, dans les années 90, les entrepôts de données sont apparus. Il s'agit d'une collection de données *orientées sujet* (organisées par thème), *intégrées* (assemblées à partir de sources hétérogènes) et *consolidées* (dans un format cohérent et normalisé), *non volatiles* (une fois une donnée entrée, elle ne change pas), *historisées* (de nouvelles données sont intégrées sans suppression des anciennes), et organisées pour le support d'un processus d'aide à la décision [67]. Leur fonctionnement est basé sur le processus ETL (« extraction, transformation, loading »), qui comporte trois étapes : (1) connexion de sources de données hétérogènes et extraction des données ; (2) transformation des données dans un format spécifique cible, basé sur des règles et des conditions ; (3) chargement dans des tables.

Ainsi, l'utilisation croissante du DPI et la mise en place de SI dans les établissements de santé, s'est accompagnée de la mise en place progressive d'entrepôts de données de santé (EDS), afin d'optimiser la réutilisation des données de santé pour la recherche, la pratique clinique, le management de l'hôpital, ou à visée administrative [61]. Les EDS permettent d'accéder à de multiples données de vie réelle issues des SI des établissements de santé [61,68–76]. Lorsque les données sont correctement qualifiées et structurées, les EDS permettent d'inclure un grand nombre de patients sans devoir revenir à leur dossier médical, et de mener des études de surveillance (suivi de populations ou de pathologies, etc.) ou observationnelles (études de cohorte, études cas-témoins, etc.) à l'échelle d'un ou de plusieurs établissements de santé [61,63,68–76].

Ces dernières années, de nombreux établissements de santé, ou regroupements d'établissements de santé, ont mis en place des EDS en France. L'hôpital européen Georges-Pompidou (HEGP) a été le premier en France à se doter en 2010 d'un EDS. D'autres exemples incluent l'Assistance Publique – Hôpitaux de

Paris (AP-HP), les Hospices Civils de Lyon, l'Assistance Publique – Hôpitaux de Marseille, le Centre Hospitalier Universitaire (CHU) Grenoble Alpes, le CHU de Rouen, le CHU de Nantes, le CHU de Toulouse, etc. Par ailleurs, certains réseaux hospitaliers travaillent à la structuration coordonnée d'EDS au sein de leurs établissements : le réseau inter-régional du Grand-Est PAISaGE, groupant les données des CHU / CHR⁹ de Nancy, Strasbourg, Metz, Besançon, Dijon et Reims ; le réseau des Hôpitaux universitaires du Grand Ouest HUGO, qui réunit l'Institut de Cancérologie de l'Ouest et les CHU de Rennes, Brest, Angers, Nantes, Tours et Orléans ; le réseau du Nord G4, qui inclut les CHU d'Amiens, Caen et Rouen. En janvier 2017, l'EDS de l'AP-HP devenait le premier EDS autorisé par la CNIL en France (autorisation du 19 janvier 2017, numéro 1980120). En 2021, la CNIL a adopté un référentiel sur les EDS reposant sur l'exercice d'une mission d'intérêt public [77].

Les infections bactériennes sévères sont couramment traitées en milieu hospitalier. Ainsi, les EDS peuvent fournir des données cliniques et microbiologiques de vie réelle, permettant la caractérisation et l'étude de patients atteints d'infections bactériennes sévères, notamment en terme d'antibiorésistance.

⁹ Centre Hospitalier Régional

2 OBJECTIFS DE LA THÈSE

« Je fabriquerai une carte du monde »

Nami, One Piece, chapitre 77

L'objectif général de ce travail de recherche est d'évaluer l'apport des données hospitalières pour l'étude des infections bactériennes résistantes aux antibiotiques et en particulier l'apport de l'EDS de l'AP-HP pour étudier les bactériémies hospitalisées.

Un premier travail concerne le processus de **construction de la base de données** clinico-microbiologiques de séjours avec une bactériémie (base BactHub). Ce travail inclut l'évaluation et la résolution des problèmes de qualité liés aux données, puis le nettoyage et la structuration des données des séjours d'inclusion, autour de la notion d'épisode de bactériémie, avec la création de variables d'intérêts.

Un travail exploratoire sur **l'apport des techniques de Traitement Automatisé du Langage Naturel (TAL) pour identification des sites primaires** des bactériémies à partir des comptes-rendus d'hospitalisation (CRH) est également présenté. Les résultats de l'extraction par TAL sont comparés à l'identification des sites primaires grâce aux codages CIM-10, et discutés.

Afin de valider les données recueillies, une **description des principales caractéristiques des patients, des épisodes de bactériémie, et des isolats bactériens** est réalisée et comparée aux données de la littérature, en particulier les données sur la résistance. Les descriptions sont effectuées selon l'origine de la bactériémie : communautaire ou nosocomiale.

Par la suite, j'ai **étudié, dans les bactériémies communautaires, le rôle de l'antibiorésistance dans la récurrence de bactériémie à 1 an**. Cette étude se focalise sur les bactériémies à *S. aureus*, *E. coli*, *Klebsiella spp.* et la résistance à la méthicilline pour *S. aureus* ou aux céphalosporines de 3^{ème} génération pour *E. coli* et *Klebsiella spp.*

Enfin, les **points clés et les défis liés à l'étude du lien entre l'exposition individuelle aux antibiotiques et l'occurrence d'infections bactériennes communautaires résistantes aux antibiotiques** sont discutés. Notamment, je mettrai en avant l'apport et les limites de la base BactHub pour répondre à cette question. Des données sur l'exposition antérieure aux antibiotiques et aux soins des patients hospitalisés avec une bactériémie étant nécessaires, je présenterai les perspectives d'enrichissement de la base BactHub avec des données issues du SNDS.

Pour retrouver votre chemin le long de cette thèse, suivez la petite noire qui s'y balade !

3 LES BACTÉRIÉMIES

3.1 Intérêt de l'étude des bactériémies

Les bactériémies sont des infections bactériennes couramment rencontrées en pratique clinique, qui s'associent à une morbi-mortalité importante (20-30% de mortalité à 30 jours) [78,79]. Elles sont couramment utilisées pour étudier l'antibiorésistance, dans de nombreux travaux de recherche mais aussi par divers réseaux de surveillance et organisations tels que l'Observatoire National de l'Épidémiologie de la Résistance Bactérienne aux Antibiotiques (ONERBA) en France ou l'ECDC en Europe [37]. En effet, le diagnostic de bactériémie s'appuie uniquement sur des critères microbiologiques, ce qui simplifie l'étape de définition des cas infectés, alors que le diagnostic d'autres infections (infections urinaires, infections respiratoires, etc) nécessiterait l'accès à des données cliniques (fièvre, symptômes, signes) et/ou radiologiques qui présentent souvent des problèmes de complétude et de qualité dans le cadre d'études à grande échelle.

3.2 Définitions

Une bactériémie correspond à la présence dans au moins un prélèvement sanguin (**hémoculture**) d'une bactérie **pathogène**, qui peut s'associer ou non à des signes cliniques, notamment une fièvre. La principale difficulté repose sur les contaminations qui sont à exclure.

Contaminations

Les contaminations sont des hémocultures positives à des bactéries non pathogènes. Elles peuvent représenter 30 à 50% des hémocultures positives, et s'associent à un surcoût important pour les patients et le système de soin, ainsi qu'à un sur-traitement et des hospitalisations injustifiées [80–82]. S'il n'y a pas de gold standard pour définir les contaminations, de nombreux travaux ont tenté de délimiter quelques règles permettant d'isoler les vrais des faux positifs, comme le type de micro-organisme identifié, le nombre d'hémocultures positives, le délai entre les hémocultures positives, et le temps de pousse (durée entre l'incubation d'un flacon d'hémoculture et sa positivation) [83]. Dans sa définition, le comité technique des infections nosocomiales et des infections liées aux soins a établi en 2007 une liste des micro-organismes contaminants potentiels : staphylocoques à coagulase négative, *Bacillus* spp. (sauf *B. anthracis*), *Corynebacterium* spp., *Propionibacterium* spp., *Micrococcus* spp., et les autres micro-organismes commensaux ou de l'environnement à potentiel pathogène comparable. Pour définir une

bactériémie à l'un de ces micro-organismes, deux hémocultures positives au même micro-organisme, prélevées lors de prélèvements différents du sang, à des moments différents, et dans un intervalle rapproché (un délai maximal de 48h est habituellement utilisé), sont exigées [84]. Dans leur définition des hémocultures, les Centres pour le contrôle et la prévention des maladies (CDC) des Etats-Unis précisent la liste des commensaux courants, potentiels contaminants : staphylocoques à coagulase négative, *Bacillus* spp. (sauf *B. anthracis*), micro-organismes diphthéroïdes (*Corynebacterium* spp., sauf *C. diphtheria*), *Propionibacterium* spp., les streptocoques du groupe viridans, *Aerococcus* spp. *Micrococcus* spp., *Rhodococcus* spp, ou tout autre commensal de la liste établie par le réseau national de sécurité en santé (National Healthcare Safety Network, NHSN) [85,86].

Bactériémies nosocomiales et communautaires

Comme toutes les infections bactériennes, les bactériémies peuvent se classer en deux groupes : les bactériémies associées aux soins et les bactériémies communautaires. Une infection est dite **associée aux soins** (IAS) si elle survient au cours ou au décours d'une prise en charge (diagnostique, thérapeutique, palliative, préventive ou éducative) d'un patient, et si elle n'était ni présente, ni en incubation au début de la prise en charge. Les IAS englobent les infections **nosocomiales** (contractées dans un hôpital) et les infections associées à des soins pratiqués en dehors de l'hôpital (soins à domicile, EHPAD, etc.). Par opposition, les infections **communautaires** désignent les infections survenues dans la communauté, en dehors de toute exposition au soin. Les bactériémies communautaires sont le plus souvent approchées en identifiant les bactériémies survenues dans les 48 heures après l'admission dans un établissement de santé [87,88]. Cette définition est la plus couramment utilisée ; elle comprend parfois l'exclusion des patients transférés d'un autre établissement de santé ou hospitalisés dans les semaines précédant l'hospitalisation pour infection [88]. Cependant, cette définition est imparfaite, puisqu'elle peut inclure à tort les patients ayant été exposés aux soins avant l'admission (par exemple, des soins infirmiers à domicile).

Dans la suite du manuscrit, « communautaire » désignera les bactériémies « community-onset », donc survenues dans les 48h après l'admission dans un établissement de santé. Au sein des bactériémies communautaires, l'on peut distinguer les bactériémies « communautaires pures » (sans exposition préalable au soin, l'équivalent de la catégorie « community-acquired ») et celles de survenue communautaire mais associées aux soins (l'équivalent de la catégorie « healthcare-associated »). Les différentes définitions des bactériémies sont présentées dans le **Tableau 3.1**, en fonction des classifications.

Tableau 3.1 Classification des bactériémies

	Exposition aux soins avant admission	Survenue <48h après l'admission
Classification internationale		
Community-onset, community-acquired	X	✓
Community-onset, healthcare-associated	✓	✓
Hospital-acquired	✓ ou X	X
Classification française		
Communautaire	X	✓
Associée aux soins	✓	✓
Associée aux soins, nosocomiale	✓ ou X	X
Manuscrit de thèse		
Communautaire, pure	X	✓
Communautaire, associée aux soins	✓	✓
Nosocomiale	✓ ou X	X

Bactériémies primaires et bactériémies secondaires

Au sein des bactériémies, on peut distinguer les bactériémies secondaires et primaires, respectivement, selon la présence ou l'absence d'une source d'infection définie en dehors du sang.

Cette source d'infection primaire (ou site d'infection primaire) peut par exemple être une infection urinaire à *E. coli*, avec passage secondaire d'*E. coli* dans le sang, causant une bactériémie secondaire. Cela peut également être un impétigo (infection cutanée) à *S. aureus*, avec passage du *S. aureus* dans le sang.

3.3 Quelques chiffres

D'après la littérature, près de la moitié des bactériémies prises en charge à l'hôpital sont communautaires, et l'autre moitié sont nosocomiales [80,89,90]. Peu de travaux évaluent la proportion des bactériémies primaires, qui semble avoisiner 12-14% [89,90].

Dans une étude multicentrique sur 24 hôpitaux en France, menée sur une période de 2 mois en 1993, Brun-Buisson *et al.* ont identifié 842 épisodes de bactériémie (55% communautaires) associés à des signes cliniques, chez 832 adultes [89]. Par ailleurs, 14% des bactériémies étaient primaires, sur la base de l'examen clinique et des caractéristiques microbiologiques des patients [89]. Les bactériémies secondaires avaient comme site d'infection primaire : une infection urinaire (21%), une infection digestive (18%), une infection respiratoire basse (16%), une infection liée à un cathéter intravasculaire (11%), une infection de la peau et des tissus mous (8%), une infection ostéoarticulaire (2%), une infection neuroméningée (2%), des sources d'infection multiples (4%), ou une autre source d'infection (1%).

Dans une étude multicentrique menée dans 3 centres hospitaliers aux États-Unis sur une période de 12 mois en 1992-1993, 843 épisodes de bactériémie ont été identifiés, dont 48% étaient communautaires [80].

Enfin, Ortega *et al.* ont identifié 3900 épisodes de bactériémie (40% communautaires) entre 2003 et 2006 dans l'unité de médecine interne du service d'urgences d'un hôpital en Espagne, après exclusion des patients neutropéniques [90]. Pour les épisodes communautaires, un site d'infection primaire a été recherché par un infectiologue spécialiste, sur la base des antécédents, de l'histoire de la maladie, de l'examen physique, et des résultats d'une analyse des urines et d'une radiographie thoracique du patient. Au total, 12% des bactériémies communautaires étaient primaires ; pour les bactériémies secondaires, les sites primaires n'étaient pas précisés [90].

4 L'ENTREPÔT DES DONNÉES DE SANTÉ DE L'AP-HP

4.1 Présentation et modalités d'accès

L'AP-HP (<https://www.aphp.fr>) est la plus grande structure hospitalo-universitaire d'Europe, avec 39 hôpitaux (22 474 lits ; **Annexe 1**) principalement situés en région parisienne, totalisant 1,5 million d'hospitalisations par an (environ 10% de l'ensemble des hospitalisations en France) [76]. Depuis 2014, une plateforme analytique a été développée sous forme d'un EDS agréant les données quotidiennes des patients hospitalisés. En septembre 2016, la commission médicale d'établissement centrale de l'AP-HP a fixé les règles d'accès et d'utilisation de son nouvel EDS [91]. En janvier 2017, l'EDS de l'AP-HP devenait le premier EDS autorisé par la CNIL en France (autorisation du 19 janvier 2017, numéro 1980120).

L'EDS de l'AP-HP contient des données démographiques, des diagnostics, des actes, des résultats de biologie et d'imagerie, des données de prescriptions de médicaments, des CRH associés à des passages à l'hôpital, des données médico-administratives relevant du PMSI, mais aussi les données de mortalité extrahospitalière issues du registre national des décès (INSEE¹⁰). Les examens d'imagerie, stockés sur un PACS (Picture Archive and Communication System), peuvent également être mis à disposition selon le projet de recherche. L'EDS de l'AP-HP dispose d'une capacité totale de stockage de 2 pétaoctets, dont plus de 150 téraoctets sont déjà utilisés, hors imagerie (<https://eds.aphp.fr/eds>). Il rassemble aujourd'hui les données de plus de 11 millions de patients traités dans les 39 établissements de l'AP-HP.

Les modalités d'accès aux données de l'EDS de l'AP-HP dépendent de la nature du projet, et du demandeur d'accès. Pour les professionnels de santé de l'AP-HP, un accès facilité aux données de leur équipe de soins est possible, par une simple demande au coordonnateur de l'EDS du groupe hospitalier. Si le porteur de projet n'est pas un professionnel de santé de l'AP-HP ou si le périmètre du projet est plus large que sa propre équipe de soins, une demande d'accès et un protocole de recherche doivent être soumis au Comité Scientifique et Ethique (CSE) de l'EDS de l'AP-HP. L'accès aux données est conditionné par l'accord du CSE, qui se réunit une fois par mois, et évalue notamment si le projet est conforme à la méthodologie de référence MR004. Telle que définie par la CNIL, la méthodologie de référence MR004 « encadre les traitements de données à caractère personnel à des fins d'étude, évaluation ou recherche n'impliquant pas la personne humaine. Il s'agit plus précisément des études ne répondant pas à la définition d'une recherche impliquant la personne humaine, en particulier les études

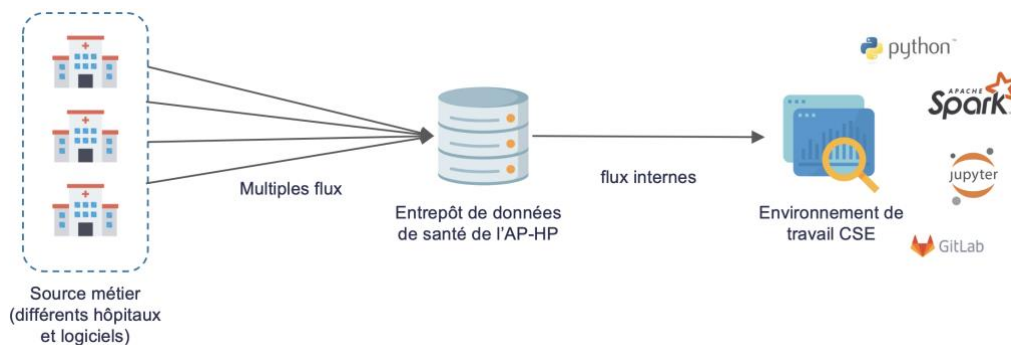
¹⁰ Institut national de la statistique et des études économiques

portant sur la réutilisation de données. La recherche doit présenter un caractère d'intérêt public. Le responsable de traitement s'engage à ne collecter que les données strictement nécessaires et pertinentes au regard des objectifs de la recherche. » [92]



Le premier accord délivré par le CSE pour un projet de recherche date d'avril 2018. Une fois un projet accepté par le CSE, les données sont dé-identifiées avant d'être mises à disposition dans un espace projet dédié avec des accès personnels pour les membres du projet, appelé « environnement de travail CSE », sur une plateforme Jupyter en ligne (**Figure 4.1**). La structure de Jupyter s'organise autour de 3 principaux éléments : l'application web, chargée d'afficher les notebooks et permettant à l'utilisateur d'interagir avec ; les notebooks, qui sont des fichiers avec une extension .ipynb, où l'utilisateur va rentrer du code, et qui afficheront les résultats générés par le lancement du code ; des kernels, chargés d'interpréter et exécuter le code contenu dans les notebooks. Selon le langage informatique utilisé dans un notebook, il peut s'agir d'un kernel R, Python, Spark¹¹, HiveQL. La liste des études acceptées par le CSE est disponible sur : <http://recherche.aphp.fr/eds/etudes>.

Figure 4.1 Les données de l'AP-HP, de la source à l'espace projet



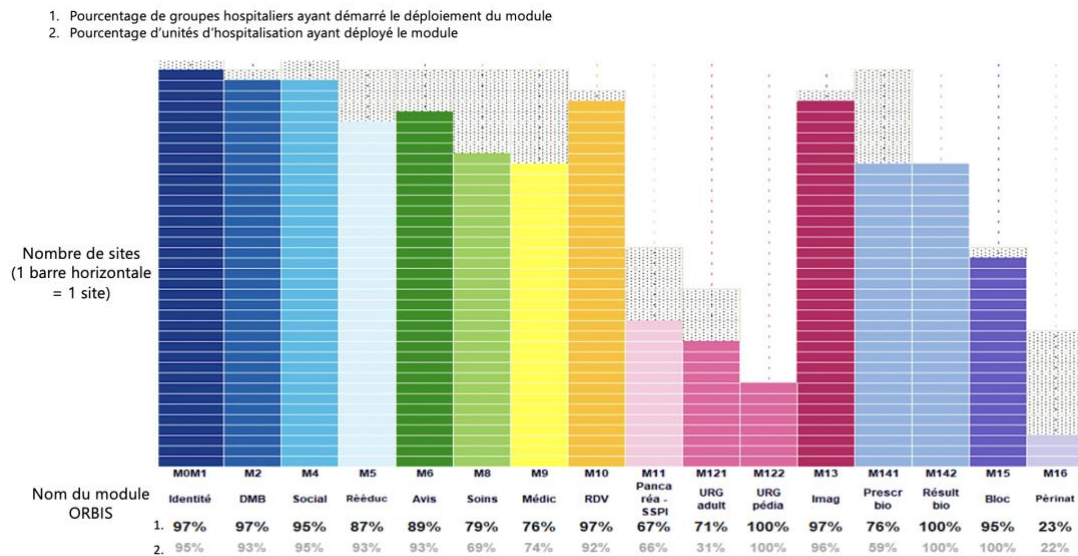
4.2 Source et intégration des données

L'EDS de l'AP-HP s'appuie principalement sur le système d'information Orbis (AGFA Healthcare), qui a été la base du déploiement du DPI commun au sein de l'AP-HP. Ce déploiement a été progressif, et a duré 9 ans, de 2010 à 2019. Depuis la pandémie Covid-19, des efforts importants ont été déployés pour informer le contenu de l'EDS et un site internet est dédié (<https://eds.aphp.fr/>). A l'initiation de cette thèse, peu d'informations sur les données et sur leur qualité étaient disponibles. Le logiciel Orbis possède de multiples modules (dossier médical standard, dossier de soins, Orbis médicaments, nutrition, dossier

¹¹ Spark (Apache Spark) n'est pas un langage de programmation, mais une infrastructure de traitement parallèle Open source développée pour analyser un grand volume de données de manière distribuée (cluster computing), implémentée dans plusieurs langages de programmation comme Python (PySpark), R (RSpark), etc.

de chirurgie ambulatoire, etc.), dont le déploiement est à ce jour hétérogène et incomplet, selon les services hospitaliers et les établissements (**Figure 4.2**) [93]. Par exemple, au 9 novembre 2021, le module M2, du dossier médical, était déployé dans 97% (38/39) des sites, et 93% des unités d'hospitalisations ; le module M121, urgences adultes, était déployé dans 71% (12/17) des sites où il doit être déployé, et 31% des unités d'hospitalisation.

Figure 4.2 État des lieux du déploiement des modules d'Orbis au 09/11/2021



Chaque couleur représente un module d'Orbis, les barres horizontales colorées correspondent aux sites hospitaliers où le module a été déployé, et les parties grisées correspondent à des sites où le module doit être déployé. Pour chaque module, les % en noir (1.) correspondent aux % de sites hospitaliers ayant démarré le déploiement d'un module donné, parmi les sites qui doivent le déployer et ceux en gris (2.) aux pourcentages d'unités d'hospitalisation ayant déployé un module donné.

D'autres logiciels sont utilisés par le personnel de l'AP-HP ; les données de ces logiciels sont incomplètement intégrées à l'EDS. Pour le circuit du médicament (prescription, dispensation, administration), plusieurs logiciels sont utilisés selon les hôpitaux et les services : Phedra, Actipidos, Orbis médicament, Chimio, Pharma, Dx CARE, et PCS. La plupart des hôpitaux utilisent Phedra et/ou Actipidos, et le déploiement du module Orbis médicaments est toujours en cours. A la date du 16/03/22, seules les données d'Orbis médicament sont disponibles dans la plateforme de l'EDS de l'AP-HP, et les données de Chimio et de Phedra sont indiquées en cours d'intégration (**Tableau 4.1**).

Ainsi, les données disponibles au sein de l'EDS de l'AP-HP s'enrichissent progressivement. Elles sont toujours, à ce jour, incomplètes. La cartographie ci-dessous (**Figure 4.3**) présente le statut d'intégration

des données à la date du 16/03/22. Les applications indiquées correspondent à celles au travers desquelles les données sont mises à disposition (Jupyter¹², i2b2¹³, Cohort360¹⁴, Pilote¹⁵) [94].

Tableau 4.1 Intégration des données relatives au circuit du médicament dans l'EDS de l'AP-HP, selon leur source, à la date du 16/03/2022 [95]

Outil	Constructeur	Description	Types de données	Statut
ORBIS	AGFA	Dossier patient informatisé déployé sur l'ensemble de hôpitaux de l'AP-HP.	Prescription médicament	Disponible
CHIMIO		Application permettant la gestion des chimiothérapies et la dispensation de médicaments dans ce cadre.	Prescription médicament	En cours d'intégration
PHEDRA		Application utilisée par la pharmacie pour la traçabilité du médicament et notamment la dispensation médicamenteuse.	Prescription médicament	En cours d'intégration
DxCare		Dossier patient informatisé déployé à l'HEGP et Beaujon.	Prescription médicament	À faire

Figure 4.3 Cartographie des données de l'EDS de l'AP-HP en date du 16/03/2022



¹² Jupyter est une application web utilisée à l'EDS pour donner accès à un espace de travail dédié à chaque projet.

¹³ L'application i2b2 (informatics for integrating biology and the bedside) est une application qui permet de sélectionner des patients dans une base de données structurée suivant le format du même nom. Elle est décommissionnée depuis 2021.

¹⁴ Cohort360 est une application développée par la Direction des systèmes d'information (DSI) de l'AP-HP, permettant de visualiser les données de l'entrepôt et de sélectionner un ensemble de patients suivant des critères pour répondre à une question de recherche.

¹⁵ Pilote est une plateforme de « business intelligence », mettant à disposition de nombreux rapports et tableaux de bord pour le pilotage de l'AP-HP, au niveau des services, des sites, des groupes hospitalo-universitaires (GHU) et global AP-HP. Ces rapports et tableaux de bord présentent des indicateurs sur trois grands thèmes : Activité/Soins, Budget/Finances, Ressources Humaines.

4.3 Information des patients

Depuis août 2017, les patients pris en charge à l'AP-HP sont informés individuellement de la réutilisation de leurs données à des fins de recherche, par une mention sur les CRH remis aux patients individuellement, et par la présence sur le site de l'EDS d'une présentation de l'entrepôt et d'une information complète sur les recherches validée par le CSE de l'AP-HP. Pour les patients n'ayant pas été soignés à l'AP-HP depuis août 2017, une information collective est prévue sur le site de l'EDS. Les patients souhaitant s'opposer à ce que des données à caractère personnel les concernant fassent l'objet de traitements automatisés ayant pour finalité la recherche peuvent le faire par un formulaire en ligne, en application de l'article 56 de la Loi Informatique et Libertés.

5 BASE DE DONNÉES BACTHUB

5.1 Points techniques

5.1.1 Entrepôt de données de santé de l'AP-HP

Le travail de structuration de la base BactHub s'est fait en collaboration étroite avec l'équipe de l'EDS de l'AP-HP (qui fournissait les données, en particulier Stéphane Bréant, Christel Daniel, Patricia Martel, Cécile Poret, Patricia Serre, Tom Denat), et avec les deux data engineers mis à disposition par le Health Data Hub (HDH) pendant 18 mois pour aider à structurer la base de données, Salma El Oualydy et Maeva Kos.

Sélection de la population

La population retenue était tous les patients majeurs avec (1) au moins un séjour avec une hémoculture positive (à toute bactérie, à l'exclusion des bactéries intracellulaires, ou spirochètes, ou à champignon de type *Candida* ou *Fusarium*, **Annexe 2**) ou un code cim-10 de bactériémie dans les diagnostic principal (DP)¹⁶, diagnostic relié (DR)¹⁷ ou diagnostic associé significatif (DAS)¹⁸ des unités d'hospitalisation du séjour (A40 streptocoque ; A41, staphylocoque, BGN, anaérobies, autres ; A32.7, *Listeria* ; A39.4, méningocoque ; A42.7, *Actinomyces* ; A02.1, *Salmonella* ; A54.8, gonocoque ; A48.0, *Clostridium*), (2) entre le 1^{er} janvier 2016 et le 31 décembre 2019, (3) dans l'un des 23 hôpitaux avec une activité de soins aigus (MCO) de l'AP-HP.

Pour les patients inclus, les données de l'ensemble de leurs séjours (séjours d'inclusion avec bactériémie, et autres séjours entre le 1^{er} janvier 2016 et le 31 décembre 2019) ont été mises à disposition en mars 2020 dans un environnement privé de travail sur la plateforme Jupyter.

Dans un deuxième temps, les données de leurs séjours survenus du 1^{er} janvier au 31 décembre 2020 ont été mises à disposition, pour permettre un suivi des patients jusque 1 an après leur séjour d'inclusion.

¹⁶ Le DP est le problème de santé qui a motivé l'admission du patient dans une unité d'hospitalisation. Le DP du résumé de sortie standardisé n'est pas disponible.

¹⁷ Le DR a pour rôle, en association avec le DP et lorsque celui-ci n'y suffit pas, de rendre compte de la prise en charge du patient en termes médico-économiques. C'est une information complétant et précisant le DP.

¹⁸ Un DAS est un diagnostic ou un autre problème de santé significatif d'une majoration de l'effort de soins et de l'utilisation des moyens, par rapport à la morbidité principale.



Processus de structuration

Les données mises à disposition s'organisaient dans 15 tables : informations patients, informations séjours, codes CIM-10¹⁹, actes de la classification commune des actes médicaux (CCAM)²⁰, observations physiologiques, résultats de laboratoire, résultats de microbiologie, résultats d'imagerie (métadonnées du PACS²¹), documents médicaux (CRH, lettres, etc.), Groupe Homogène de Malades²², liste des passages dans des unités d'hospitalisation au cours d'un séjour, données médicamenteuses, liste des concepts et items des ontologies, liste des fournisseurs de données.

Ces tables contiennent l'ensemble des données incluses dans l'EDS pour les patients retenus. Ainsi, outre les séjours d'intérêt, l'ensemble des séjours antérieurs et postérieurs sans limite de date sont aussi mis à disposition. Le numéro d'identification du patient et celui du séjour sont les clés de chaînage, c'est-à-dire les variables permettant de lier les tables entre elles.

L'absence d'un dictionnaire de données à jour ou de documentation (données médicamenteuses issues du logiciel Phedra) pour expliquer le contenu des tables a rendu leur exploration difficile. Ainsi, une première étape a consisté à explorer les différentes tables et mettre en place un dictionnaire.

Par la suite, deux processus différents ont été menés en parallèle : (1) qualification²³ et qualité des données ; (2) structuration autour des épisodes de bactériémie.

Au fil de la qualification des données et de la remontée de nouvelles données sources dans l'EDS depuis les hôpitaux de l'AP-HP, de multiples versions de la base ont été livrées successivement par l'EDS. Ces versions permettaient de corriger des problèmes de qualité identifiées dans les bases précédentes, ou d'apporter de nouvelles données nécessaires à l'étude. Par exemple, les données médicamenteuses étaient partiellement disponibles et se basaient uniquement sur le module ORBIS médicament. Pour les enrichir, plusieurs livraisons de données issues du logiciel Phedra eurent lieu, la dernière en juin 2021.

Au total, la construction de la base BactHub a nécessité 21 livraisons de données sur une période de 15 mois.

¹⁹ La CIM-10 est la 10e révision de la Classification statistique internationale des maladies et des problèmes de santé connexes, une liste de classification médicale établie par l'Organisation mondiale de la santé.

²⁰ La CCAM est une nomenclature de la Sécurité sociale française regroupant le codage des gestes pratiqués par les médecins, les chirurgiens-dentistes et les sage-femmes.

²¹ Picture archiving and communication system, il s'agit du système de stockage des images médicales.

²² Le groupe homogène de malades (GHM) est la catégorie élémentaire de la classification médico-économique propre au PMSI en MCO et odontologie. Tout séjour aboutit dans un seul groupe de la classification GHM selon un algorithme qui se fonde sur les informations médico-administratives du résumé de sortie standardisé.

²³ Procédés visant à améliorer la fiabilité et la cohérence des informations dans une base de données, tels que le nettoyage, le dédoublonnage, l'enrichissement des données, etc.

Qualification et qualité des données

Les données livrées présentait de nombreux problèmes de qualité. La première étape de leur résolution était d'identifier si la source avait pour origine les flux de données remontées dans l'EDS (depuis la saisie des données à l'hôpital jusqu'aux tables reçues). Si tel était le cas, une nouvelle livraison avait lieu. Dans le cas contraire, les problèmes étaient résolus par la mise en place de règles de traitement. Afin d'optimiser le processus de résolution et de suivi des problèmes de qualité, un Gitlab²⁴ spécifique a été mis en place par l'équipe de l'EDS.

Les problèmes rencontrés étaient principalement liés à (1) **des discordances** entre (i) les numéros de patient des tables (numéros de patient dans certaines tables, non présents dans la table « patient » qui doit contenir tous les patients sélectionnés), (ii) les numéros de séjours des tables (numéros de séjours dans certaines tables, non présents dans la table « visite » qui doit contenir tous les séjours retenus), (iii) les numéros de séjours de la table contenant les données de microbiologie, et celle contenant les données de séjour, (iv) les différentes dates du séjour (par exemple : dates d'observations physiologiques non incluses dans la période de la visite correspondante, etc.) ; (2) **des problèmes de chaînage** des données de microbiologie issues du logiciel GLIMS²⁵ et des séjours hospitaliers correspondants ; (3) **des doublons** (~30% des séjours) ; (4) **des données manquantes** : (i) sur le passage dans les unités de l'hôpital (63% des séjours), (ii) sur la date de fin de séjour (~35%) ; (iii) sur les observations biologiques (~30%), (iv) sur les services hospitaliers dans lesquels sont rédigés les CRH (5) **des données aberrantes** sur des dates, des valeurs d'observations physiologiques (par exemple, une pression artérielle négative). Pour les données médicamenteuses issues du logiciel Phedra, un nombre important de variables avait un taux de remplissage faible (pouvant aller de 3 à 70%), les numéros de séjours de la table Phedra ne correspondaient pas avec ceux des autres tables.

Pour chaque table d'intérêt, l'évaluation de la qualité a été formalisée par un rapport évaluant : le respect des clés de chaînage, la complétude des données, leur conformité aux nomenclatures de référence et aux types des variables, leur plausibilité temporelle et atemporelle.

²⁴ GitLab est une plateforme de développement collaborative open source (société Gitlab). Se basant sur les fonctionnalités du logiciel Git, elle permet de déposer des codes sources et gérer leurs différentes versions..

²⁵ GLIMS (CliniSys) est un système d'information de laboratoire qui permet d'organiser de manière flexible et d'automatiser les processus des laboratoires. GLIMS est utilisé dans les hôpitaux de l'AP-HP. A date de la construction de la base BactHub, 12 versions différentes de GLIMS étaient utilisées dans les hôpitaux de l'AP-HP.

Structuration autour des épisodes de bactériémie

Après l'étape de qualification permettant de résoudre un nombre conséquent des écueils, les séjours d'inclusion avec bactériémie ont été identifiés, et leurs données ont été nettoyées et structurées autour de la notion d'**épisode de bactériémie**, avec la création de variables d'intérêts (processus de structuration en **Annexe 3**, nomenclatures en **Annexe 4**).

Au total, ce processus a nécessité la création de plus de 100 tables, et la mise en place de 6 algorithmes complexes pour (1) détecter les imbrications de séjours ; (2) réaliser le dédoublonnage des séjours ; (3) détecter les épisodes de bactériémie ; (4) classer les épisodes en communautaire ou nosocomial ; (5) identifier les contaminants ; (6) identifier les pires variables physiologiques et biologiques d'un épisode (+/-24h, +/-48h, +/-72h autour du début de l'épisode, ou durant le séjour) pour calculer un score de sévérité lié à l'infection, le score SOFA²⁶.

Les principales définitions permettant d'identifier les épisodes de bactériémie, de les classer en communautaire ou nosocomial, et d'exclure les contaminations d'hémocultures, sont présentées ci-dessous.

Épisode de bactériémie :

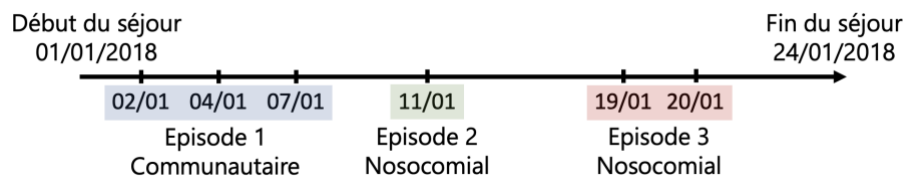
- 1) Identifié à partir des données de microbiologie : présence d'une ou de plusieurs hémocultures positives, avec 72h ou moins d'écart entre chaque hémoculture positive et la précédente. Ainsi, un même patient peut présenter plusieurs épisodes de bactériémies par séjour.
- 2) Identifié à partir des codes CIM-10 : les séjours sans hémoculture positive, mais qui avaient un code CIM-10 de bactériémie en DP, DR, ou DAS ont été considérés comme des épisodes de bactériémie. Dans ce cadre, un séjour correspond à un épisode.

Dans la suite de ce manuscrit, ce que j'appellerai épisode de bactériémie correspondra, sauf mention contraire, à un épisode identifié à partir des données de microbiologie.

Épisode de bactériémie communautaire ou nosocomial : un épisode de bactériémie est classé en communautaire si (1) sa première hémoculture positive est prélevée moins de 48h après le début du séjour hospitalier et (2) le patient n'est ni transféré d'un établissement de santé ni sorti d'un établissement de l'AP-HP dans les 7 jours précédant son séjour d'inclusion. Dans le cas contraire, l'épisode est classé en nosocomial. Ainsi, un même patient peut présenter un seul épisode communautaire, mais plusieurs épisodes nosocomiaux, par séjour (**Figure 5.1**). Par définition, seuls les épisodes définis à partir des données de microbiologie ont pu être classés en communautaire ou nosocomial.

²⁶ Sepsis-related Organ Failure Assessment score

Figure 5.1 Exemple d'un séjour, avec plusieurs épisodes de bactériémie



Chaque date notée correspond à une hémoculture positive.

Exclusion des contaminations d'hémocultures :

Les espèces considérées comme potentiellement contaminantes étaient *Corynebacterium sp*, *Staphylococcus epidermidis*, *Staphylococcus pettenkoferi*, *Staphylococcus haemolyticus*, *Staphylococcus hominis*, *Staphylococci à coagulase négative*, *Staphylococcus capitis*, *Staphylococcus hominis subsp. hominis*, *Micrococcus luteus*, *Staphylococcus capitis subsp. capitis*, ou *Staphylococcus warneri*. Ce choix s'est appuyé sur les espèces classiquement listées comme contaminantes, et sur la prévalence des bactéries dans la base (les espèces potentiellement contaminantes mais rares ont été ignorées).

Au sein d'un même épisode, défini à partir des données de microbiologie, nous avons considéré qu'une hémoculture à l'une de ces espèces n'était pas une contamination si les 3 conditions suivantes étaient réunies : (1) présence d'au moins une autre hémoculture positive à la même espèce dans les 72h ; (2) présence pour les deux hémocultures d'au moins 6 résultats de sensibilité aux antibiotiques sur 8 disponibles pour les antibiotiques suivants : oxacilline, céfoxitine, tobramycine, gentamicine, rifampicine, érythromycine, clindamycine, et cotrimoxazole ; (3) similarité entre les deux hémocultures des résultats des tests de sensibilité disponibles pour ces antibiotiques. A noter, les résultats de sensibilité des bactéries aux antibiotiques dans la base reposaient sur les valeurs critiques définies par le Comité de l'Antibiogramme de la Société Française de Microbiologie (CA-SFM) en association avec l'European Committee on Antimicrobial Susceptibility Testing (EUCAST), et la catégorisation qualitative en « sensible » (S), « intermédiaire » (I), et « résistant » (R) [96].



Écueils

Bien que très riche en données, la base finale présente pour certaines variables des taux de remplissage faibles, comme celles concernant les traitements antibiotiques prescrits pendant le séjour. En effet, seuls 46 % des séjours avec bactériémies avaient des données médicamenteuses de prise d'antibiotiques. Or, ce taux devrait être proche de 100% si les cas de contaminations ont bien été identifiés. Ce faible taux pourrait s'expliquer par une remontée incomplète des données Phedra et ORBIS médicament, et/ou l'absence de remontée de données d'autres logiciels de prescriptions comme Actipidos. Ce faible taux de remplissage concernait aussi les données physiologiques et biologiques. Ces données sont importantes dans le cadre des bactériémies car elles permettent de calculer le score de sévérité SOFA.

Pour les bactériémies communautaires, les taux de remplissage des variables nécessaires pour le calcul du score SOFA sont présentés dans le **Tableau 5.1** Ces paramètres étant systématiquement évalués lors d'un passage en réanimation, ces taux ont été estimés sur l'ensemble des séjours et sur ceux avec un passage en réanimation.

Tableau 5.1 Valeurs minimales et maximales des taux de remplissage (%) des variables physiologiques et biologiques nécessaires pour le calcul du score de SOFA sur l'ensemble des séjours avec bactériémie communautaire et sur ceux avec un passage en réanimation - BactHub 2016-2019

	Ensemble des séjours N = 33 385	Séjours avec passage en réanimation N = 12 707
Variables physiologiques		
Diurèse	16-34	24-49
Fraction inspirée d'oxygène	7-9	15-20
Score de Glasgow	10-32	12-29
Pression artérielle moyenne	34-66	37-68
Variables biologiques		
Pression artérielle d'oxygène	30-54	37-77
Taux de plaquettes	58-93	56-94
Taux de bilirubine totale	53-88	53-92
Taux de créatinine	58-93	56-94

Les taux de remplissage sont fonction de la variable et de l'année. Pour toutes les variables, le taux minimal est en 2016 et le taux maximal en 2019. Comme attendu, les taux s'améliorent au fil des années, et sont plus élevés dans les séjours avec passage en réanimation. Cependant, ils restent globalement faibles, ne permettant pas de calculer un score SOFA.

Enfin, les données microbiologiques n'incluent pas de données moléculaires de génotypage de la résistance aux antibiotiques et de typage des souches. Ainsi, les bactéries peuvent être analysées à l'échelle de l'espèce, mais pas des souches, et leurs résultats d'antibiorésistance reposent sur la catégorisation qualitative en S, I, R.

Comme présenté dans la partie II.4., un autre point essentiel dans la classification et la caractérisation des bactériémies est la détermination de leur caractère primaire ou secondaire, selon la présence ou l'absence d'un site d'infection primaire en dehors du sang. Plusieurs manières permettraient de déterminer les sites d'infection primaires : (1) l'établissement d'un ensemble de critères cliniques, microbiologiques, et radiologiques pour chaque type d'infection bactérienne existante, puis l'identification de ces infections pour les patients de la base grâce à ces règles. Cette option se heurte au nombre important d'infections bactériennes à définir, et à la présence incomplète de certaines données nécessaires à leur diagnostic. Par exemple, la présence ou l'absence de fièvre est une donnée indispensable pour différencier une infection urinaire basse (cystite) d'une infection urinaire haute (pyélonéphrite) chez la femme ; (2) l'utilisation des codages cim10, qui est la manière couramment

utilisée pour identifier les infections bactériennes dans les bases de données médico-administratives [29,97,98], mais dont la fiabilité n'est pas démontrée dans le cas de la recherche des sites primaires de bactériémie ; (3) la recherche des sites d'infection primaire directement dans les CRH des patients par TAL, méthode qui n'a à ce jour pas été explorée dans ce cadre d'application.

Par la suite, nous nous sommes appuyés sur les codages CIM-10 pour définir le site infectieux, néanmoins une analyse exploratoire utilisant le TAL a été réalisée et est présenté ci-dessous.

5.1.2 Enrichissement par Traitement Automatisé du Langage : analyse exploratoire

Ce travail a été réalisé en collaboration étroite avec Xavier Tannier, expert en TAL (Professeur à Sorbonne Université, chercheur au Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, UMRS_1142 Inserm). Après avoir défini de manière conjointe les besoins du projet et la stratégie pour y répondre, Xavier Tannier a implémenté les programmes informatiques nécessaires, et j'ai établi le corpus terminologique et évalué l'algorithme sur un échantillon de CRH. Ce travail a nécessité de nombreux échanges et itérations pour arriver au résultat présenté ici.

Bref rappel : le TAL dans le domaine biomédical

Le DPI étant avant tout un outil informant et traçant les soins, et non un outil de recueil de données pour la recherche, la plupart des données contenues dans les DPI et par extension les EDS restent à ce jour narratives et non structurées. Pour élargir le potentiel de ces données pour la recherche, en parallèle de l'apparition du DPI, les techniques de TAL dans le champ de la santé se sont développées depuis les années 60 [99]. Le TAL regroupe « l'ensemble des recherches et développements visant à modéliser et reproduire, à l'aide de machines, la capacité humaine à produire et à comprendre des énoncés linguistiques dans des buts de communication » [100]. L'adjectif naturel fait référence au langage humain, par opposition au langage formel (C, etc.). Le TAL est par essence une discipline multidisciplinaire, faisant le lien entre les théories de la linguistique notamment du sous-langage, de l'informatique, de la représentation du discours, et -dans le cas du domaine biomédical- les expertises en biologie, en médecine et en santé.

Le langage humain peut être décrit par plusieurs niveaux : le vocabulaire, appelé *lexique* ; la forme des mots, appelée *morphologie* ; la grammaire, appelée *syntaxe* ; le sens, appelé *sémantique* ; l'influence du contexte sur la signification, appelée *pragmatique* ; la relation entre les énoncés d'un écrit, appelée *cohérence* ; les paragraphes et documents, appelés *discours* [99,101]. Du fait de la complexité de ces

niveaux et de leur grande variabilité selon les contextes et les champs d'études, le TAL est un vrai défi. Dans le domaine biomédical, la linguistique est différente de celle utilisée dans le langage général, avec un style plus télégraphique, un lexique différent, et l'absence de respect de la morphologie des mots et de certaines règles de grammaire : on parle de « sous-langage » [102,103]. La sémantique, la cohérence et le discours sont également différents. Par exemple, l'expression « secondaire à » indique une relation de causalité qui n'est pas existante dans le langage général [101].

Deux approches principales de TAL existent. La première s'appuie sur la mise en place manuelle d'une terminologie et de règles qui permettent à l'ordinateur d'analyser et d'interpréter le texte. **La terminologie** consiste en un lexique, listant pour les concepts recherchés toutes les manières dont ils peuvent se décliner dans le langage clinique (par exemple, pour le concept *Infection*, cela peut être *infection neuro-méningée, infection neuroméningée, infection du cathéter, endocardite*, etc.). Certains des termes dans le lexique peuvent être fixes (sans variation possible dans une lettre ou dans le nombre, on parle d'*entité*), soumis à une certaine marge d'erreurs (par exemple, *lésion* pourra être identifié si le terme utilisé est *lésions*), ou correspondre à des termes à exclure, ou des expressions régulières. **Les règles** spécifient l'interaction des différents concepts les uns avec les autres au sein de phrases et du discours (par exemple, supprimer les déclinaisons du concept *Infection* s'il y a un concept de négation avant, comme dans *absence d'argument pour une infection*). Les règles peuvent spécifier et s'appuyer sur des patterns sémantiques, des règles de syntaxe, et d'éventuelles relations hiérarchiques ou non, propres au champ d'étude (par exemple traitement-traite-maladie, infection-site-bactérie-porte d'entrée) [101]. En terme informatique, les règles se traduisent par des formalismes symboliques ou logiques (par exemple, les expressions régulières) ou des formalismes statistiques [101]. Cette approche est très laborieuse, puisqu'elle requiert la mise en place et la mise à jour régulière d'une terminologie et de règles détaillées couvrant toutes les déclinaisons possibles des concepts recherchés et les règles de syntaxe et d'organisation du discours permettant de les identifier.

Dans les années 2000, une approche, basée sur **l'apprentissage automatique supervisé**, s'est développée dans le domaine biomédical [99,101]. Cette approche peut permettre de classifier ou ordonner des items, prédire une information, reconnaître des entités ou des structures sémantiques [99]. De mise en œuvre plus complexe, les algorithmes associés apprennent sur des données et des tâches déjà résolues par l'humain (entraînement supervisé) à reproduire et généraliser les décisions pour les appliquer à des données nouvelles.

Méthodologie d'extraction des concepts

Dans le cadre de notre travail exploratoire, à la suite de l'étude de 20 CRH annotés sur papier, l'approche basée sur l'établissement d'une terminologie et de règles semblait être suffisante, plus simple à mettre en œuvre et a donc été retenue.

Supposons ce texte, dans un CRH :

A J1 de l'admission, bactériémie à *S. aureus* sur érysipèle du membre inférieur, d'évolution favorable sous Augmentin. Absence d'argument en faveur d'une endocardite. A J20, pyélonéphrite obstructive compliquée d'un choc septique à *E. coli* BLSE. Drainage urinaire et antibiothérapie, évolution favorable. Passage en médecine interne le 13 avril. Le 20 avril, fièvre sans porte d'entrée. Hémoculture positive à *E. cloacae*, avec différentielle de pousse en faveur d'une infection sur cathéter.

Le patient a présenté trois épisodes infectieux : (1) une bactériémie communautaire à *S. aureus* de porte d'entrée primaire cutanée, avec élimination d'une suspicion d'endocardite (2) une infection urinaire nosocomiale à *E. coli* (3) une bactériémie nosocomiale à *E. cloacae* de porte d'entrée primaire une infection sur cathéter. Notre objectif était de pouvoir identifier les portes d'entrées primaires de chaque bactériémie. Donc, l'algorithme de TAL doit pouvoir : (1) identifier séparément chaque bactériémie et sa bactérie, (2) identifier le site primaire lié à chaque bactériémie, (3) ignorer les informations liées à l'infection urinaire, qui n'est pas une bactériémie, et (4) ignorer la suspicion d'endocardite.

A cette fin, il était nécessaire d'établir un corpus de référence incluant (1) une terminologie listant pour les concepts à extraire les termes possibles (par exemple, *S. aureus*, *Staphylococcus aureus*, *Staph aureus*, Staph doré, etc. sont des possibilités de la bactérie *Staphylococcus aureus*) ; (2) une terminologie listant des termes de négation (exemple : *absence d'argument en faveur*, *l'a écarté*, *a écarté*, *ne ressemblent pas*, *ne révèlent pas*, etc.) ; (3) des règles d'identification des concepts (par exemple, identifier une infection uniquement s'il y a dans la même phrase une notion de bactériémie ou d'une bactérie, elle-même liée ailleurs dans le CRH à une bactériémie).

L'établissement du corpus de référence a nécessité 11 itérations, comprenant chacune successivement (1) l'établissement d'une première version du corpus ; (2) l'extraction de concepts à partir de ce corpus, sur un nombre croissant de CRH, allant de 50 à 200 ; (3) la correction du corpus en fonction des erreurs identifiées. Ce processus s'est appuyé sur **brat**, un outil basé sur le Web pour l'ajout de notes à des documents textuels existants [104,105].

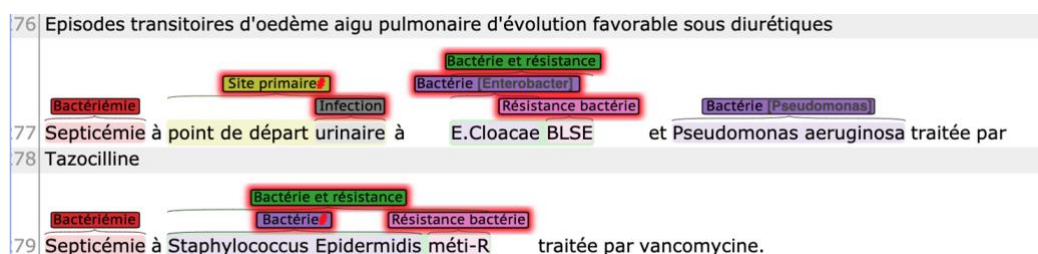
Le corpus final comprenait : une terminologie de 1753 termes, ainsi que 271 termes de négation, et 15 règles d'identification des concepts. Parmi ces règles, beaucoup s'appuyaient sur le périmètre de la phrase, qui est par définition imparfait. En effet, il est très fréquent dans les CRH étudiés que les phrases ne se terminent pas par un point, ou qu'il y ait des sauts de ligne au milieu d'une phrase. Ceci complique la détection des frontières de la phrase par les algorithmes de segmentation automatique, et génère des erreurs. Par ailleurs, certaines règles se basaient sur la délimitation de sections au sein des CRH (en-tête, motif de séjour, antécédents, histoire de la maladie, etc.), grâce à un outil mis en place par un membre de l'EDS de l'AP-HP. Ces contraintes de sections définissent pour chaque concept, dans quelle(s) section(s) il doit être identifié. Par exemple, si l'identification d'une bactérie, en dehors de la section examens complémentaires, histoire de la maladie, évolution médicale, conclusion, ou la section autre, suppression de la bactérie.

En plus de ce corpus de référence, nous devons fixer pour l'algorithme de TAL un seuil d'acceptation d'erreurs, c'est-à-dire un seuil d'acceptation de variations morphologiques par rapport aux termes de la terminologie²⁷. En effet, nous souhaitons que l'algorithme identifie l'infection si « érysipèle » était noté dans un CRH au lieu de « érysipèle ». Plus le seuil choisi était proche de 1, moins on autorisait l'algorithme à s'écarter de la terminologie ; plus il était proche de 0, plus on autorisait les erreurs. Nous avons extrait les informations de plusieurs dizaines de CRH avec plusieurs seuils (de 0,7 à 0,95). La comparaison des résultats nous a permis de fixer le seuil à 0,8.

La **Figure 5.2** montre un exemple de CRH annoté grâce au TAL. Grâce au corpus mis en place, on voit que l'algorithme de TAL a bien identifié les concepts *Bactériémie*, *Site primaire*, *Enterobacter*, *Pseudomonas*, et *Résistance*.



Figure 5.2 Pré-annotation de concepts dans un document médical de la base BactHub, visualisée sur l'outil brat



²⁷ Nous utilisons l'algorithme simstring pour détecter ces variations (<https://pypi.org/project/simstring-pure/>). Cet algorithme s'appuie sur la proportion de trigrammes de caractères communs entre deux termes.

Structuration et liaison au reste de la base

L'étape finale consistait à structurer dans des tables de données pouvant être reliées au reste de la base, les concepts identifiés par l'algorithme. Par ailleurs, en présence de plusieurs bactériémies, il était nécessaire d'identifier les concepts correspondants à chacune d'entre elles. Par exemple, dans la **Figure 5.2**, il paraissait important que le *point de départ urinaire* soit en relation avec la bactériémie à *E. cloacae* et *Pseudomonas*, et non pas la bactériémie à *Staphylococcus epidermidis*. Deux tables principales ont donc été structurées : la première incluant tous les concepts identifiés avec le détail des termes correspondants dans chaque CRH, un identifiant unique, et le numéro de séjour des patients ; la seconde recréant toutes les relations entre les concepts liés à une même bactériémie.

Pour chaque concept, tous les termes identifiés dans la base en lien avec ce concept ont été normalisés et catégorisés (par exemple, pour le concept *Infection*, j'ai classé les termes *pyélonéphrite*, *pyélo*, *prostatite* dans la catégorie *Appareil_urinaire*, et les termes *médiastinite*, *endocardite*, *El du cœur*, *infection du médiastin*, dans la catégorie *Cardio_vasculaire*). Les numéros de séjours ont permis de chaîner les deux tables à la base BactHub.

5.2 Contrôle qualité

5.2.1 Données issues de l'entrepôt de données de santé de l'AP-HP

Afin d'éviter les biais de recrutement, et s'assurer de la stabilité temporelle des données, nous avons évalué la qualité de la remontée des données de microbiologie dans les 23 hôpitaux inclus. Pour chaque hôpital et année, nous avons calculé le ratio du nombre de séjours avec un code CIM-10 de bactériémie mais sans hémoculture positive, divisé par le nombre de séjours avec une bactériémie en microbiologie. Un hôpital était exclu si (1) pour une année donnée, ce ratio était supérieur à 0,25 ou si (2) le nombre de séjours avec une bactériémie en microbiologie était très faible. A l'issue de cette étape, le nombre d'hôpitaux inclus a été restreint à 14 hôpitaux avec une activité de soins aigus, représentant plus de 75% des lits de soins aigus au sein de l'AP-HP (Tableau 5.2), et couvrant environ 22% des séjours de soins aigus dans la région île de France [106].

Tableau 5.2 Hôpitaux avec une activité de soins aigus à l'AP-HP

Hôpital	Type d'activité	Inclus	Nombre de lits
Ambroise-Paré	A ^a	Oui	497
Antoine-Béclère	A	Non ^c	411
Armand-Trousseau	A	Non ^{c,d}	~70 (adultes)
Avicenne	A	Oui	543
Beaujon	A	Oui	394
Bicêtre	B ^b	Oui	1,027
Bichat-Claude Bernard	B	Oui	916
Cochin	A	Oui	1,074
Georges Pompidou	A	Non ^d	827
Henri-Mondor	A	Non ^{c,d}	843
Hôtel-Dieu	A	Non ^d	44
Jean-Verdier	A	Oui	300
Lariboisière et Fernand-Widal	B	Oui	976
Louis-Mourier	B	Non ^c	494
Necker-Enfants Malades	A	Non ^{c,d}	~200 (adultes)
Paul-Brousse	B	Oui	771
Pitié-Salpêtrière	B	Oui	1,717
Raymond-Poincaré	B	Non ^c	362
Robert-Debré	A	Non ^d	~60 (adultes)
Saint-Antoine	A	Oui	679
Saint-Louis	A	Oui	700
Tenon	A	Oui	559

^aA : soins aigus ; ^bB : activité mixte de soins aigus et de soins de longue durée et réadaptation ; ^cexclusion en raison d'un ratio > 0.25 ; ^dexclusion car faible remontée des données microbiologiques.

Par ailleurs, les séjours sans hospitalisation complète, ou n'ayant pas passé au moins une nuit à l'hôpital, ont été exclus.

5.2.2 Données issues du programme de Traitement Automatisé du Langage

Afin d'évaluer le programme informatique de TAL, nous avons comparé ses extractions pour 300 épisodes de bactériémie à une lecture experte des CRH. Puis, nous avons discuté la place de cette technique pour l'analyse de la base BactHub en (1) évaluant la valeur informative des CRH de la base BactHub, selon la présence ou non d'une information sur les bactériémies identifiées en microbiologie, puis en (2) comparant les résultats du programme de TAL et ceux des codages CIM-10.

Évaluation du programme informatique de TAL

Pour évaluer les extractions effectuées par le programme de TAL, la référence est la lecture experte des comptes-rendus, pour l'identification des bactériémies-bactéries, et des sites primaires. 300 épisodes

avec bactériémie en microbiologie ont été tirés au sort et l'ensemble des CRH (n = 966) liés aux séjours de ces épisodes ont été retenus. En effet, la date des CRH et leur ordre de rédaction au fil du séjour n'étant pas disponibles, il n'était pas possible d'identifier uniquement les CRH de fin de séjour. Pour chaque CRH j'ai regardé si une bactériémie était mentionnée, et si oui, quelle(s) étai(en)t la(les) bactérie(s) et les éventuels sites primaires. Ces résultats ont été comparés à ceux identifiés par le TAL.

Pour estimer les qualités du programme de TAL, nous sommes placés dans la situation la plus pénalisante, correspondant le mieux à la réalité. Ainsi une information partiellement correcte était inexacte. Cela peut concerner deux cas de figure : (1) une information incomplète (faux négatif) et (2) trop d'informations incluant l'information attendue (faux positif). La répartition des vrais positifs (VP), faux positifs (FP), vrais négatifs (VN) et faux négatifs (FN) pour l'évaluation du programme de TAL est présentée dans le **Tableau 5.3.a et 5.3.b.1**. Les détails de la classification sont présentés en **Annexe 5**.

La performance du TAL a été évaluée par deux indices : le rappel ou sensibilité, et la précision. Soient VP, FP, VN et FN, le Rappel $R = VP / (VP + FN)$, la Précision $P = VP / (VP + FP)$. Par exemple, dans le cas de l'identification des bactériémies-bactéries, plus le rappel est proche de 1, meilleure est la part des bactériémies-bactéries identifiés, parmi la totalité des bactériémies-bactéries. Plus la précision est proche de 1, plus élevé est le taux de bactériémies-bactéries réelles parmi celles identifiées.

Notre objectif était d'avoir un score de Rappel proche de 1 pour l'identification des bactériémies-bactéries, et un score de Précision proche de 1 pour les sites primaires. En effet, dans la vraie vie et donc en l'absence de lecture experte des CRH, la référence pour les bactériémies-bactéries est la microbiologie, qui est disponible dans la base BactHub. Il est donc possible d'éliminer les bactériémies-bactéries FP en recoupant les identifications du programme de TAL avec les données de microbiologie. Néanmoins, il n'y a pas de référence pour les sites primaires, les identifications doivent donc être les plus proches possibles de la réalité.

Le Rappel et la Précision pour l'identification des bactériémies-bactéries, et pour l'identification des sites primaires, étaient respectivement de 0,90 - 0,86 et de 0,70 - 0,88. Ces valeurs étaient > 0,85 et donc satisfaisantes mis à part le Rappel pour les sites primaires. Ainsi, bien que la part des vrais sites primaires soit élevée parmi ceux identifiés (Précision > 0,85), il y a de nombreux sites primaires non identifiés (Rappel = 0,70).

Afin d'améliorer ces scores, il faut minimiser les FP et les FN. Un retour aux CRH a permis d'identifier la(les) raison(s) de ces FP ou FN (exemples et répartition en **Annexe 5**). L'identification de bactériémies-bactéries FP était surtout liée (1) à l'existence de multiples infections autres que des bactériémies dans les CRH, (2) à la quasi-impossibilité de segmentation de l'information au sein des phrases, et (3) aux

erreurs de segmentation des CRH. Les bactériémies-bactéries FN étaient principalement dues aux erreurs de segmentation des CRH et d'identification des fins de phrase, et à la complexité de mise en place d'une terminologie exhaustive. Les causes de FP et de FN pour les sites primaires étaient superposables à celles des bactériémies-bactéries, avec en plus la présence prépondérante de situations diagnostiques complexes nécessitant une lecture experte pour déterminer le site primaire.

Si une partie de ces causes peut être améliorée (segmentation des CRH, identification des fins de phrase, enrichissement de la terminologie), d'autres semblent difficilement adressables avec l'approche terminologie-règles (segmentation de l'information au sein d'une phrase, situations diagnostiques nécessitant une lecture experte). Une possibilité serait l'utilisation d'une technique d'apprentissage automatique à type de reconnaissance d'entité, qui nécessiterait l'annotation manuelle d'un nombre important de documents, puis l'entraînement d'un modèle de reconnaissance de séquences de mots.

Tableau 5.3 Identification des bactériémies et des sites primaires dans les CRH d'épisodes de bactériémie par le programme de TAL ou les codages CIM-10, en comparaison avec une lecture experte

	Lecture experte	
	Classés +	Classés -
Performance du programme de TAL		
a. Identification de bactériémies-bactéries dans les CRH de 300 épisodes		
Classés +	186	30
Classés -	20	64
b. Identification de sites primaires		
b.1. Dans les CRH de 300 épisodes*		
Classés +	150	20
Classés -	63	67
b.2. Dans les CRH de 300 épisodes**		
Classés +	148	21
Classés -	131	0
b.3. Dans les CRH de 216 épisodes ^{#,**}		
Classés +	148	18
Classés -	50	0
Performance des codages CIM-10		
c. Identification de sites primaires pour 216 épisodes [#]		
Classés +	82	102
Classés -	32	0

*Référence pour les bactériémies-bactéries = lecture experte. **Référence pour les bactériémies-bactéries = microbiologie. #Les 216 épisodes ont des CRH avec une information complète sur les bactériémies-bactéries.

Apport du TAL en comparaison aux codages CIM-10 pour l'identification des sites primaires dans la base BactHub

Après avoir évalué le programme de TAL, nous avons étudié son apport potentiel pour la détermination des sites primaires des épisodes de la base BactHub en situation réelle, donc avec la microbiologie comme référence pour les bactériémies-bactéries, et sans retour aux CRH.

Le programme de TAL analyse des CRH, il faut donc évaluer la part d'épisodes de la base avec des CRH informatifs pour l'étude des bactériémies. J'ai comparé mon identification de bactériémies-bactéries dans les CRH des 300 épisodes tirés au sort, aux bactériémies-bactéries identifiées en microbiologie pour ces mêmes épisodes. L'information sur les bactériémies-bactéries était complète dans les CRH de 216 épisodes sur 300 (72%). Ce pourcentage est faible et représente une première limite à l'utilisation du programme de TAL pour l'identification des sites primaires dans la base BactHub. Les informations manquantes peuvent être dues à la non-intégration de certains CRH à l'EDS de l'AP-HP, ou à l'absence de mention des bactériémies dans les CRH par les soignants.

En nous limitant aux bactériémies-bactéries identifiées en microbiologie donc présentes dans la base BactHub, le Rappel et la Précision pour l'identification des sites primaires par le programme de TAL étaient de 0,53 – 0,88 pour les 300 épisodes étudiés, et 0,75 – 0,89 pour les 216 épisodes avec des CRH informatifs (**Tableau 5.3.b.2 et 5.3.b.3, Annexe 6**).

Les codages CIM-10 étant classiquement utilisés dans notre cadre d'application, nous avons comparé les résultats de ces codages à une analyse experte, pour les 216 épisodes avec une information complète disponible sur les bactériémies-bactéries (**Tableau 5.3.c, Annexe 6**) : le Rappel et la Précision pour l'identification des sites primaires étaient de 0,72 – 0,45.

Si tous les CRH de la base BactHub étaient informatifs, les Rappels du programme de TAL et des codages CIM-10 seraient comparables ; dans la situation actuelle, les codages CIM-10 ont un Rappel 1,4 fois plus élevé. En comparaison au programme de TAL, les codages CIM-10 identifiaient plus de sites primaires, mais avec plus de faux positifs. Ainsi, le programme de TAL a une Précision près de 2 fois plus élevée que les codages CIM-10. Les patients avec bactériémie présentent fréquemment de multiples infections non bactériémiantes lors de leurs séjours, qui peuvent être prolongés. Les codages CIM-10 font nécessairement référence à toutes (ou au moins une partie de) ces infections, ce qui explique leur faible Précision. Pour visualiser l'impact de ces différences sur les résultats, le **Tableau 5.4** présente la répartition des sites primaires identifiés avec le TAL, les codes CIM-10, et la lecture experte, pour les 216 épisodes avec une information complète disponible sur les bactériémies-bactéries.

Les résultats montrent une sur-représentation des bactériémies primaires (absence de site primaire) dans les identifications par le programme de TAL (du fait de nombreux FN donc d'un faible Rappel), et une sur-représentation des infections multiples dans les identifications basées sur les codes CIM-10 (du fait de nombreux FP donc d'une faible Précision). Pour le reste des sites primaires, la répartition est proche entre les techniques, avec une variabilité plus ou moins importante.

Au total, nous avons mis en place un programme de TAL basé sur l'approche terminologie-règles, pour identifier les sites primaires infectieux des bactériémies. Les extractions de bactériémies-bactéries et de sites primaires par le programme de TAL étaient correctes dans la majorité des cas. Néanmoins, l'utilisation du programme pour analyser la base BactHub se heurte à deux limites importantes, la présence de nombreux FN de sites primaires, et l'absence d'information sur les bactériémies-bactéries dans les CRH pour plus d'un épisode sur 4. En conséquence, près de la moitié des épisodes analysés n'avaient pas de site primaire identifié par le programme de TAL, soit près de 20% de plus par rapport à une lecture experte.

Si le développement croissant du TAL dans le domaine biomédical est prometteur, c'est en le couplant à l'optimisation des ressources (exhaustivité des CRH) et de leur contenu (minimisation de l'hétérogénéité, exhaustivité de l'information) que l'on pourra exploiter tout son potentiel. Par exemple, une transposition et structuration automatique des examens complémentaires dans les CRH, ou la mise en place de correcteurs automatiques, pourraient améliorer le contenu des CRH. Enfin, les soignants devraient être mieux informés du potentiel et des techniques informatiques de réutilisation des données des CRH.



Tableau 5.4 Répartition des sites primaires pour 216 épisodes de bactériémie, selon la technique d'identification des sites primaires, en pourcentage (%)

Site primaire	TAL	CIM-10	CRH
Bactériémie primaire (Absence)	49,54	25,46	30,56
Urinaire	12,04	15,28	18,06
Digestif	7,41	5,56	12,04
Infection sur matériel	5,09	3,24	11,11
Multiples	10,65	35,19	9,72
Respiratoire bas	5,56	5,56	6,02
Peau et tissus mous	3,70	1,85	6,02
Cardio-vasculaire	4,17	2,31	3,24
Ostéoarticulaire	1,39	1,39	2,78
Autres	0,46	4,17	0,46

6 CARACTÉRISTIQUES, ANTIBIORÉSISTANCE, MORTALITÉ ET INCIDENCE DES BACTÉRIÉMIES

Les caractéristiques des patients inclus dans la base BactHub, de leurs premiers épisodes de bactériémie, et de leurs isolats bactériens sont décrits. L'incidence des bactériémies sur 3 ans (2017-2019) est estimée. La représentativité et l'apport de la base de données pour la recherche sont aussi discutés.

6.1 Matériels et méthodes

Population d'étude

Les séjours de patients majeurs avec un épisode de bactériémie prouvée en microbiologie (hors contaminants) entre le 1^{er} janvier 2016 et le 31 décembre 2019 inclus dans la base BactHub ont été identifiés. Seul le premier épisode de bactériémie de chaque patient sur la période d'étude a été retenu et analysé, qu'il soit communautaire ou nosocomial.

Données recueillies

Pour chaque patient, le sexe, l'âge, les comorbidités (déterminées à partir des codes CIM-10), la date du décès (le cas échéant), le nombre d'épisodes sur la période de l'étude, ont été recueillis. Pour chaque premier épisode, les dates d'admission et de sortie, le passage en soins intensifs, les dates de prélèvement des hémocultures, les bactéries identifiées en microbiologie et leurs résultats de sensibilité aux antibiotiques ont été collectés.

Variables étudiées

Les caractéristiques des patients comprenaient le sexe (masculin/féminin), et l'âge, considéré en 5 classes : 18-35 ans, 35-50 ans, 50-65 ans, 65-80 ans, > 80 ans. Les comorbidités suivantes étaient étudiées : cancer, tumeur solide métastatique, insuffisance cardiaque, diabète, maladies vasculaires, rénales, hépatiques, systémiques, maladies pulmonaires chroniques, démence, hémiplégie. Les comorbidités, et l'incidence de comorbidité de Charlson étaient déterminés selon les travaux de Quan et al (**Annexe 4**) [107,108]. Les mortalités intra-hospitalière, intra-hospitalière si passage en soins intensifs, intra-hospitalière sans passage en soins intensifs, à 30 jours et à 90 jours ont été estimées.

Les caractéristiques des épisodes comprenaient leur caractère monomicrobien ou polymicrobien, selon le nombre d'espèces distinctes isolées, indépendamment de leur antibiorésistance. La durée de l'épisode a été calculée, de la première hémoculture positive à la fin du séjour. Le passage en soins intensifs durant

le séjour lié à l'épisode a été considéré. Les sites primaires des épisodes ont été déterminés grâce à la catégorisation des codes CIM-10 d'infection bactérienne (**Annexe 4**) [29]. Les sites primaires suivants étaient possibles : absence de site primaire, infection urinaire, infection pulmonaire, infection digestive, infection cardio-vasculaire, infection ostéoarticulaire, infection de la peau et des tissus mous, infection sur matériel, infections multiples, autres infections. Selon l'absence ou la présence d'un site primaire, l'épisode était classé en primaire ou secondaire, respectivement.

Les caractéristiques des isolats bactériens ont été analysées après déduplication par espèce et par épisode, en favorisant les résultats de sensibilité les plus résistants. Les taux d'antibiorésistance ont été calculés en suivant les combinaisons résistance-antibiotiques de l'ECDC (**Annexe 7**) [37]. Par exemple, les isolats d'*E. coli* étaient considérés résistants aux C3G s'ils étaient résistants à la céfotaxime, à la ceftriaxone ou au ceftazidime.

Analyses statistiques

Les patients, les premiers épisodes et leurs isolats bactériens ont été analysés séparément, stratifiés par catégorie de bactériémie : communautaire vs nosocomiale. Les caractéristiques des patients, des épisodes et des bactéries sont exprimées en médiane [premier quartile (Q1), troisième quartile (Q3)] pour les variables continues, et en nombre (%) pour les variables catégorielles. Les données sont rapportées pour l'ensemble de la période d'étude de 4 ans. Les données des patients sont présentées par année dans l'**Annexe 8**.

Les taux de résistance ont été comparés aux données de l'ensemble des pays de l'ECDC et aux données pour la France, dont la représentativité serait élevée. Les données de l'ECDC sont basées sur les isolats de sang ou de liquide céphalorachidien [18].

L'incidence annuelle a été calculée de 2017 à 2019 et exprimée en nombre total de séjours avec bactériémie pour 1 000 séjours. Le nombre total de séjours a été recueilli dans les données de la Statistique annuelle des établissements de santé [109]. Les séjours incidents ont été sélectionnés dans la base BactHub en excluant les séjours ayant déjà fait l'objet d'une hospitalisation avec bactériémie au cours des 12 derniers mois. L'évolution de l'incidence a été évaluée par un test de χ^2 pour la tendance, et une valeur p de $<0,05$ a été considérée comme significative.

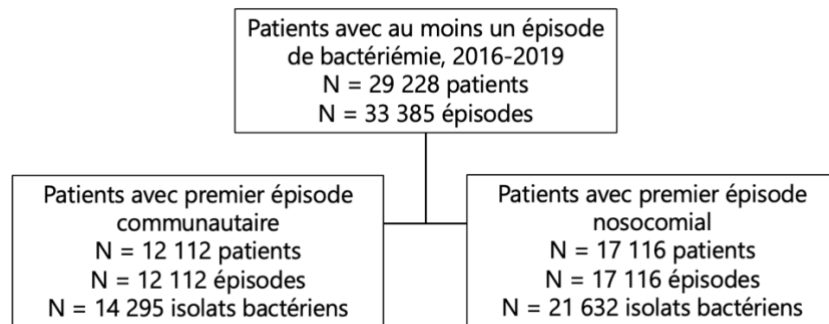
HiveQL, Python 3, PySpark 2.4.3 ont été utilisés pour effectuer les analyses statistiques.

6.2 Résultats

De 2016 à 2019, nous avons identifié 29 228 patients avec 33 385 épisodes de bactériémie (**Figure 6.1**). Leur premier épisode était communautaire pour 12 112 patients (41%), et nosocomial pour 17 116

patients (59%). La proportion de patients ayant présenté plus d'un épisode sur la période de l'étude était de 9,4% et 11,6% parmi ceux avec un premier épisode communautaire ou nosocomial, respectivement. Seuls les premiers épisodes ont été inclus.

Figure 6.1 Identification des patients avec un épisode de bactériémie et inclusion des données des premiers épisodes, BactHub 2016-2019, groupe communautaire et nosocomial



Caractéristiques des patients

Les patients avec un premier épisode communautaire étaient moins souvent des hommes (55% vs 62%) et avaient un âge médian légèrement plus élevé (68 ans vs 64 ans), avec une proportion plus importante de patients âgés (>80 ans, 23% vs 13%) (**Tableau 6.1**). Le score de Charlson médian était de 2 [1^{er} quartile- 3^{ème} quartile : 0-3] dans le groupe communautaire, et 2 [1-5] dans le groupe nosocomial. Il y avait deux fois plus de patients avec un score de Charlson nul dans le groupe communautaire (39% vs 22%). Un pourcentage important de patients avait un cancer, presque deux fois plus dans le groupe nosocomial (43% vs 25%). Un patient sur 5 présentait un diabète dans les deux groupes, avec une même proportion de diabète compliqué (~8%). L'insuffisance cardiaque (communautaire 15%, nosocomial 19%), les maladies vasculaires (communautaire 12%, nosocomial 19%) et les tumeurs solides métastatiques (communautaire 10%, nosocomiale 16%) étaient fréquentes, surtout dans le groupe nosocomial. Les mortalités intra-hospitalière (communautaire 14%, nosocomial 19%) et à 30 jours (communautaire 15%, nosocomial 18%) étaient proches, et plus importantes dans le groupe nosocomial. La mortalité à 30 jours (18%) était légèrement inférieure dans le groupe nosocomial à la mortalité intra-hospitalière (19%), possiblement du fait de la présence de séjours de durée supérieure à 30 jours dans ce groupe. La mortalité à 90 jours était de 20% dans le groupe communautaire et 26% dans le groupe nosocomial. Les patients avec un passage en soins intensifs avaient des mortalités intra-hospitalière, à 30 jours et à 90 jours élevées, à 26-28%, 24-25% et 31-32% dans les deux groupes. A contrario, les patients sans passage en soins intensifs avaient des mortalités intra-hospitalière et à 30 jours nettement plus faibles, à 8-13% et 10-13% dans les groupe communautaire et nosocomial, respectivement. Leur mortalité à 90 jours augmentait, et atteignait 15% dans le groupe communautaire, et 22% dans le groupe nosocomial. Les caractéristiques des patients étaient stables entre 2016 et 2019 (**Annexe 8**).

Tableau 6.1 Caractéristiques des patients selon le groupe, base BactHub, 2016 – 2019

	Communautaire	Nosocomial
Patients, N (%)	12 112 (41,4)	17 116 (58,7)
Sexe masculin, N (%)	6 678 (55,1)	10 576 (61,8)
Age (années), médiane [Q1-Q3]	68 [54 - 80]	64 [52 - 74]
18-35	953 (7,9)	1473 (8,6)
35-50	1 496 (12,4)	2453 (14,3)
50-65	3 002 (24,8)	5312 (31,0)
65-80	3 823 (31,6)	5607 (32,8)
>80	2 838 (23,4)	2271 (13,3)
Score de Charlson*	2 [0 - 3]	2 [1 - 5]
0	4 159 (39,1)	3 261 (22,2)
1-2	3 418 (32,2)	5 208 (35,4)
3-4	1 331 (12,5)	2 404 (16,4)
5-6	520 (4,9)	1 059 (7,2)
7-8	870 (8,2)	1 948 (13,3)
>8	330 (3,1)	822 (5,6)
Comorbidités*, N (%)		
Cancer	2 709 (25,5)	6 264 (42,6)
Tumeur solide métastatique	1 054 (9,9)	2 363 (16,1)
Insuffisance cardiaque	1 572 (14,8)	2 815 (19,2)
Diabète	2 102 (19,8)	2 940 (20,0)
Non compliqué	1 315 (12,4)	1764 (12,0)
Complication chronique	787 (7,4)	1176 (8,0)
Maladie vasculaire	1 261 (11,9)	2 716 (18,5)
Coronaire	185 (1,7)	496 (3,4)
Périphérique	597 (5,6)	1 237 (8,4)
Cérébrovasculaire	581 (5,5)	1 347 (9,2)
Maladie rénale	1 337 (12,6)	2 152 (14,6)
Maladie hépatique	1 004 (9,5)	1 757 (11,9)
Légère	659 (6,2)	945 (6,4)
Modérée / sévère	345 (3,3)	812 (5,5)
Maladie pulmonaire chronique	716 (6,7)	1 177 (8,0)
Démence	532 (5,0)	562 (3,8)
Paralysie (hémiplégie / paraplégie)	296 (2,8)	941 (6,4)
Maladie systémique	154 (1,5)	240 (1,6)
Mortalité, N (%)		
Intra-hospitalière	1 636 (13,5)	3302 (19,3)
Avec passage en soins intensifs	993 / 3 758 (26,4)	2 110 / 7 615 (27,7)
Sans passage en soins intensifs	643 / 8 354 (7,7)	1 192 / 9 501 (12,6)
À 30 jours	1 765 (14,6)	3 028 (17,7)
Avec passage en soins intensifs	951 / 3 758 (25,3)	1 788 / 7 615 (23,5)
Sans passage en soins intensifs	814 / 8 354 (9,7)	1 240 / 9 501 (13,1)
À 90 jours	2 376 (19,6)	4 451 (26,0)
Avec passage en soins intensifs	1 154 / 3 758 (30,7)	2 411 / 7 615 (31,7)
Sans passage en soins intensifs	1 222 / 8 354 (14,6)	2 040 / 9 501 (21,5)

* Données disponibles pour 10 628 et 14 702 patients des groupes communautaire et nosocomial.

Caractéristiques des épisodes

Une minorité d'épisodes étaient polymicrobiens (**Tableau 6.2**, communautaires 14%, nosocomiaux 19%). Les épisodes nosocomiaux avaient une durée médiane deux fois plus élevée que les épisodes communautaires (15 jours vs 8 jours). La moitié des épisodes communautaires avaient une durée inférieure à une semaine, et seuls 10% d'entre eux duraient plus d'un mois. La durée des épisodes nosocomiaux se répartissait de manière quasi-équivalente dans les catégories : ≤ 7 jours, 7-14 jours, 14-30 jours, et >30 jours. Près d'un épisode communautaire sur trois et un épisode nosocomial sur deux étaient marqués par un passage en soins intensifs.

Les 10 sites primaires les plus souvent identifiés sont présentés dans le **Tableau 6.2**. Aucun site primaire n'était retrouvé dans 22% des épisodes communautaires, et 25% des épisodes nosocomiaux. Des sites d'infection multiples étaient fréquemment identifiés, en particulier dans les épisodes nosocomiaux (35% contre 24%). Dans le cas des bactériémies secondaires, les principales infections étaient les infections urinaires (épisodes communautaires 19%, nosocomiaux 9%), les infections respiratoires basses (communautaires 11%, nosocomiaux 10%) et les infections digestives (communautaires 9%, nosocomiaux 6%).

Tableau 6.2 Caractéristiques des épisodes de bactériémie selon le groupe, base BactHub, 2016 – 2019

	Communautaire	Nosocomial
Épisodes, N (%)	12 112 (41,4)	17 116 (58,7)
Polymicrobien	1 660 (13,7)	3 318 (19,4)
Durée médiane (jours)	8,0 [3,0 - 17,0]	15,0 [7,0 - 29,0]
≤ 7	5 903 (48,7)	4 592 (26,8)
7-14	2 599 (21,5)	3 908 (22,8)
14-30	2 350 (19,4)	4 663 (27,2)
30-60	992 (8,2)	2 669 (15,6)
>60	268 (2,2)	1 284 (7,5)
Passage en soins intensifs, N (%)	3 758 (31,0)	7 615 (44,5)
Site d'infection primaire*, N (%)		
Bactériémie primaire [#]	2292 (21,6)	3606 (24,5)
Bactériémie secondaire		
Infections multiples	2544 (23,9)	5202 (35,4)
Infection urinaire	2034 (19,1)	1252 (8,5)
Infection respiratoire basse	1151 (10,8)	1511 (10,3)
Infection digestive	968 (9,1)	839 (5,7)
Infection sur matériel	594 (5,6)	1054 (7,2)
Infection de la peau et des tissus mous	386 (3,6)	493 (3,4)
Infection cardio-vasculaire	178 (1,7)	107 (0,7)
Infection ostéoarticulaire	155 (1,5)	119 (0,8)
Autre	109 (1,0)	316 (2,2)

* Données disponibles pour 10 628 et 14 702 épisodes communautaires et nosocomiaux. # Absence de site primaire.

Caractéristiques des isolats bactériens et antibiorésistance

Après dédoublement par espèce et par épisode, 35 927 isolats bactériens ont été analysés. Les dix espèces bactériennes les plus identifiées sont présentées dans le **Tableau 6.3**. Les trois espèces les plus souvent isolées étaient *E. coli*, *S. aureus* et *K. pneumoniae* dans les deux groupes. *E. coli* était identifié dans un épisode communautaire sur quatre (27%), soit presque deux fois plus que dans les épisodes nosocomiaux (15%). *S. aureus* et *K. pneumoniae* étaient identifiés à des taux comparables dans les deux groupes (11-12% et 5-7% respectivement). Les quatrièmes et cinquièmes espèces les plus identifiées étaient *Streptococcus pneumoniae* (4%) et *Enterococcus faecalis* (3%) dans les épisodes communautaires ; *Pseudomonas aeruginosa* (7%) et *Staphylococcus epidermidis* (7 %) dans les épisodes nosocomiaux.



Tableau 6.3 Distribution des isolats bactériens selon le groupe, base BactHub, 2016 – 2019

	Communautaire	Nosocomial
Isolats bactériens uniques*, N (%)	14 295 (39,8)	21 632 (60,2)
<i>Escherichia coli</i>	3 832 (26,8)	3 151 (14,6)
<i>Staphylococcus aureus</i>	1 672 (11,7)	2 513 (11,6)
<i>Klebsiella pneumoniae</i>	754 (5,3)	1 500 (6,9)
<i>Streptococcus pneumoniae</i>	529 (3,7)	-
<i>Staphylococcus epidermidis</i>	354 (2,5)	1 403 (6,5)
<i>Pseudomonas aeruginosa</i>	345 (2,4)	1 427 (6,6)
<i>Enterococcus faecalis</i>	448 (3,1)	1 069 (4,9)
<i>Proteus mirabilis</i>	225 (1,6)	-
<i>Streptococcus agalactiae</i>	215 (1,5)	-
<i>Streptococcus pyogenes</i>	192 (1,3)	-
<i>Enterobacter cloacae</i>	-	832 (3,9)
<i>Enterococcus faecium</i>	-	627 (2,9)
<i>Bacteroides fragilis</i>	-	346 (1,6)
<i>Staphylococcus haemolyticus</i>	-	320 (1,5)

Les taux de résistance des 5 bactéries les plus identifiées sont présentés dans le **Tableau 6.4**. Ces taux étaient plus importants dans les épisodes nosocomiaux, avec une différence marquée pour *K. pneumoniae* et une différence plus faible pour *E. coli* et *S. aureus*. La résistance d'*E. coli* aux C3G concernait 13-17% des isolats, et la résistance aux fluoroquinolones 22-23% des isolats. La résistance de *K. pneumoniae* aux C3G concernait 21% des épisodes communautaires et 37% des épisodes nosocomiaux, et la résistance aux fluoroquinolones 21% des épisodes communautaires et 33% des épisodes nosocomiaux. La résistance aux carbapénèmes restait rare pour *E. coli* (0,1-0,5%), mais atteignait 2% dans les épisodes nosocomiaux à *K. pneumoniae*. Les isolats de *S. aureus* étaient résistants à la méticilline dans 11% des épisodes communautaires et 14% des épisodes nosocomiaux. Leur résistance aux fluoroquinolones était de 10-12%, alors que la résistance à la rifampicine restait rare (1-2%). La résistance à la pénicilline des isolats communautaires de *S. pneumoniae* était de 19%, et 5% étaient résistants aux

fluoroquinolones. La plupart des isolats nosocomiaux de *S. epidermidis* étaient résistants à la méticilline (85%) ou aux quinolones (71%), et un isolat sur six était résistant à la rifampicine (16%). Un isolat nosocomial de *P. aeruginosa* sur cinq était résistant aux carbapénèmes (20%), 17% étaient résistants à la pipéracilline-tazobactam, 11% étaient résistants à la ceftazidime et 12% aux fluoroquinolones. Enfin, la résistance à la gentamicine à haut niveau concernait 8% des isolats communautaires d'*E. faecalis*, et celle à la vancomycine 0,2% des isolats.

Les taux de résistance aux C3G (C3G-R) des bactériémies à *E. coli* et *K. pneumoniae*, et les taux de SARM, ont été analysés selon le site d'infection primaire et le type d'épisode (**Figure 6.2**). Seuls les sites primaires avec au moins 50 isolats par espèce ont été considérés (*E. coli* et *K. pneumoniae* : infection digestive, infection urinaire, bactériémie primaire, infections multiples ; *S. aureus* : bactériémie primaire, infection respiratoire basse, infection sur matériel, infection urinaire, infections multiples, infection de la peau et des tissus mous). Le taux d'*E. coli* C3G-R et de *K. pneumoniae* C3G-R était proche dans les épisodes communautaires avec infection digestive (9% et 15%, respectivement), sans site primaire (11%, 14%), et avec infections multiples (16%, 21%). Ce taux était deux fois plus élevé dans les épisodes communautaires à *K. pneumoniae* de site primaire urinaire (31 % vs 14%). La résistance aux C3G était plus fréquente dans les épisodes nosocomiaux, et fluctuait entre 13% et 19% pour *E. coli* selon le site primaire, et entre 27% et 44% pour *K. pneumoniae* selon le site primaire. Pour *S. aureus*, le taux de résistance à la méticilline dans les épisodes communautaires était de 10-12% quand il n'y avait pas de site primaire, ou dans les infections respiratoires basses, sur matériel, ou multiples. Il atteignait 18% dans les infections de la peau et des tissus mous, et 22% dans les infections urinaires. Le taux de SARM parmi les isolats nosocomiaux de *S. aureus* était similaire aux épisodes communautaires sans site primaire ou avec infection sur matériel (10-11%), et plus élevé dans les infections respiratoires basses (16%), les infections multiples (17%), et les infections urinaires (25%).

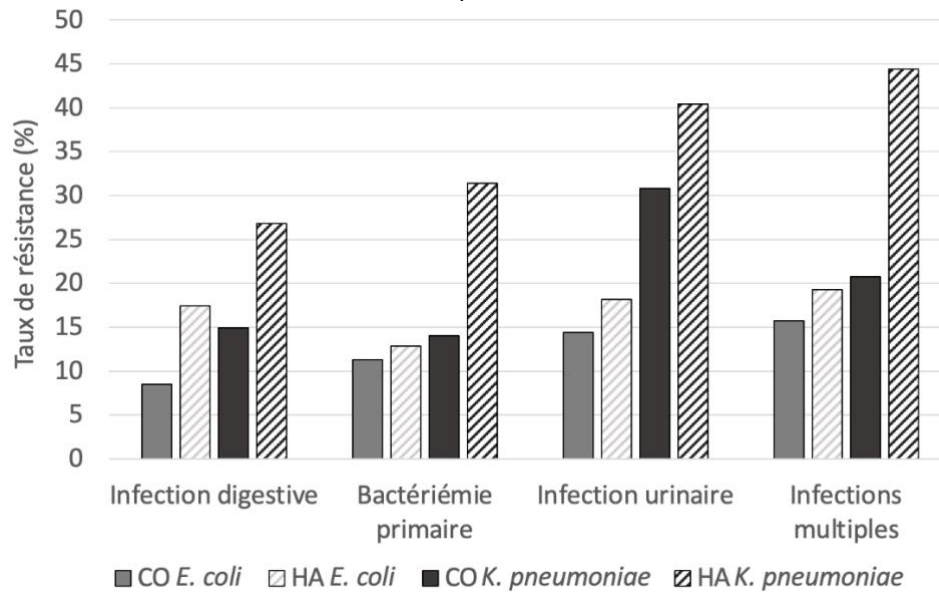
Tableau 6.4 Principaux taux de résistance aux antibiotiques (en pourcentage) des isolats bactériens selon le groupe, base BactHub, 2016 – 2019

	Communautaire	Nosocomial
<i>Escherichia coli</i>		
Aminopénicillines	62,0	68,6
C3G	13,1	17,2
Carbapénèmes	0,1	0,4
Fluoroquinolones	21,9	23,0
Aminoglycosides	10,6	12,9
<i>Staphylococcus aureus</i>		
SARM	11,2	14,2
Rifampicine	1,4	1,7
Fluoroquinolones	10,1	12,3
<i>Klebsiella pneumoniae</i>		
C3G	21,0	37,1
Carbapénèmes	0,4	1,9
Fluoroquinolones	21,6	33,2
Aminoglycosides	15,3	27,1
<i>Streptococcus pneumoniae</i>		
Pénicilline*	18,7	-
Macrolides	19,0	-
Fluoroquinolones	4,8	-
C3G	0,9	-
<i>Staphylococcus epidermidis</i>		
Méticilline	-	84,8
Rifampicine	-	16,2
Fluoroquinolones	-	70,7
<i>Pseudomonas aeruginosa</i>		
Pipéracilline-tazobactam	-	16,7
Ceftazidime	-	11,2
Carbapénèmes	-	20,4
Fluoroquinolones	-	12,1
Aminoglycosides**	-	6,8
<i>Enterococcus faecalis</i>		
Gentamycine (haut niveau)	8,1	-
Vancomycine	0,2	-

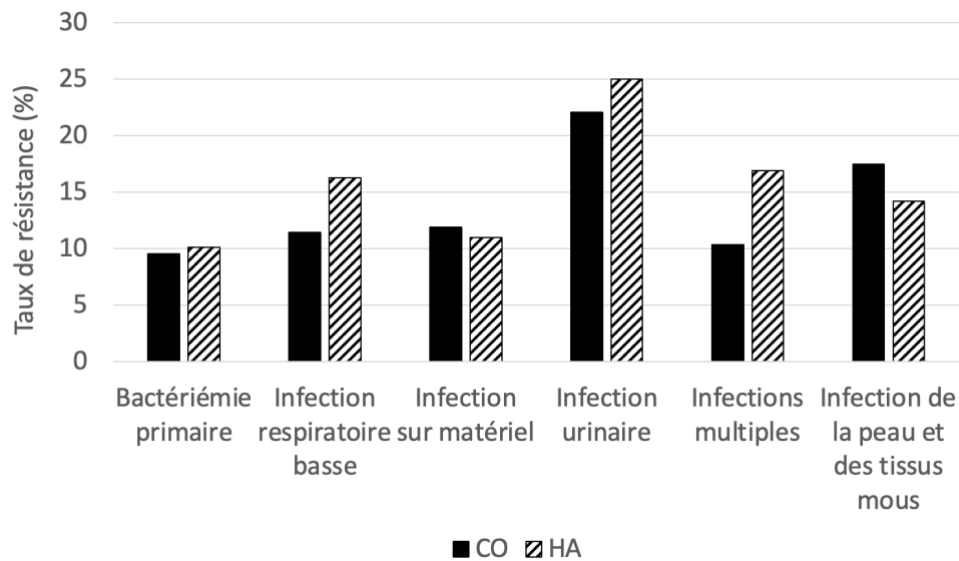
Abréviations : C3G : céphalosporines de 3^{ème} génération ; SARM : *S. aureus* résistant à la méticilline. Pour tous les antibiotiques testés, sauf la gentamycine pour *E. faecalis*, les souches intermédiaires étaient considérées résistantes. * La priorité a été donnée aux résultats de sensibilité à la pénicilline plutôt qu'à ceux de l'oxacilline. ** Résistance à la tobramycine.

Figure 6.2 Principaux taux de résistance aux antibiotiques des isolats bactériens dans le sang par site d'infection primaire et selon le groupe, base BactHub, 2016-2019

A. Résistance aux C3G des isolats d'*E. coli* et de *K. pneumoniae*



B. Résistance à la méticilline des isolats de *S. aureus*



Comparaison avec les données de l'ECDC

Dans l'ensemble, les taux de résistance d'*E. coli* étaient plus élevés que ceux rapportés à l'ECDC par la France pour toutes les résistances évaluées (C3G, 15% vs 9-11% ; fluoroquinolones, 22% vs 16-17% ; aminoglycosides, 12% vs 7-8%), et étaient proches des taux globaux de l'ECDC (**Annexe 9**) [18,37]. Le taux de SARM (13%) et les taux de résistance de *K. pneumoniae* (C3G, 32% ; fluoroquinolones, 29% ; aminoglycosides, 23%) étaient proches de ceux de la France. Les taux de résistance de *P. aeruginosa* (pipéracilline-tazobactam, 15% ; fluoroquinolones, 12% ; ceftazidime, 11%) étaient similaires à ceux de la

France, sauf pour les carbapénèmes (20% vs 13-16%) et les aminoglycosides (7% vs 8-11%). Les isolats d'*E. faecalis* présentant une résistance de haut niveau à la gentamicine (9 % vs 12-15 %) et les isolats de *S. pneumoniae* présentant une résistance à la pénicilline (21 % vs 25 %) étaient moins fréquents que dans les données de la France.

Incidence

Après exclusion des séjours ayant déjà fait l'objet d'une hospitalisation avec bactériémie sur les 12 derniers mois, 364 021 séjours ont été retenus en 2017, 359 534 en 2018 et 351 367 en 2019. Parmi eux, 7 288 étaient associés à un épisode de bactériémie en 2017, 7 531 en 2018, et 7 767 en 2019. Les taux d'incidence annuels étaient donc de 20,0 en 2017 (IC95 19,6-20,5), 20,9 en 2018 (IC95 20,5-21,4), et 22,1 séjours avec bactériémie pour 1 000 séjours en 2019 (IC95 21,6-22,6). L'incidence augmentait significativement de 2017 à 2019 ($p < 0,0001$).



6.3 Discussion

La proportion des épisodes communautaires était de 41%, et celle des épisodes nosocomiaux de 59%, ce qui concordait avec la littérature [80,89,90]. Parmi les épisodes, 17% étaient polymicrobiens ; cette proportion semblait deux fois plus élevée que celle d'anciens travaux (~7-9%) [80,89,110,111]. Les mortalités intra-hospitalière et à 30 jours étaient de 14-15% pour les patients du groupe communautaire, et de 18-19% pour ceux du groupe nosocomial. La mortalité intra-hospitalière augmentait nettement chez les patients avec passage en soins intensifs, chez qui elle atteignait 26-28% dans les deux groupes. A contrario, les patients sans passage en soins intensifs avaient une mortalité intra-hospitalière de 8% et 13%, dans les groupes communautaire et nosocomial. La mortalité à 30 jours était similaire dans le groupe communautaire, mais inférieure dans le groupe nosocomial, comparée aux résultats de deux études danoises datant d'une dizaine d'années (mortalité à 30 jours des épisodes nosocomiaux, 25-35%) [78,79]. Peu d'études ont évalué la mortalité à 90 jours en cas de bactériémie [112,113]. Elles retrouvaient une mortalité de 6-7% pour les bactériémies non compliquées à entérobactéries [113], 32% pour les bactériémies hospitalisées à *S. aureus* [112]. A notre connaissance, aucune étude n'a estimé la mortalité à 90 jours des patients avec bactériémie, indépendamment de l'espèce bactérienne. Dans notre travail, les taux de mortalité à 90 jours variaient de 20% pour les patients du groupe communautaire à 26% pour ceux du groupe nosocomial. La mortalité à 90 jours atteignait 31-32% chez les patients ayant été admis en soins intensifs pendant leur hospitalisation, alors que les patients sans passage en soins intensifs avaient une mortalité à 90 jours plus faible, de 15% dans le groupe communautaire, et 22% dans le groupe nosocomial. Ces taux globalement élevés pourraient s'expliquer en partie par la proportion importante de patients avec des cancers et des maladies cardiovasculaires, qui étaient les causes

principales de mortalité au long terme dans les bactériémies dans une étude de Nielsen et al [114]. Dans une étude, Jensen et al. ont montré que la récurrence de bactériémie était un facteur de risque indépendant de mortalité à 1 an [111]. Au total, près d'un patient sur dix a présenté plus d'un épisode sur la période de l'étude, ce qui pourrait avoir augmenté les taux de mortalité [111].

Dans l'ensemble, les taux d'antibiorésistance dans la base de données Bacthub étaient globalement proches des taux rapportés par la France auprès de l'ECDC pour *K. pneumoniae*, *P. aeruginosa*, *S. aureus* et *E. faecalis* [18,37]. Le taux de résistance à la pénicilline de *S. pneumoniae* était plus faible dans la base BactHub. A contrario, les taux de résistance de *P. aeruginosa* aux carbapénèmes et d'*E. coli* étaient nettement plus élevés dans notre base. D'après les données régionales recueillies dans le cadre des missions PRIMO, pour la résistance aux C3G, et SPARES pour la production d'une BLSE chez *E. coli*, l'île de France faisait partie des régions avec les taux de résistance les plus élevés sur la période de l'étude [51,52,54,55]. Pour les autres résistances d'*E. coli*, il n'existe pas de données régionales. Bien que les données de la mission PRIMO se limitent aux prélèvements urinaires, ces résultats suggèrent une circulation plus importante de souches d'*E. coli* résistantes aux antibiotiques dans la région île de France, qui pourrait partiellement expliquer nos résultats. La résistance de *P. aeruginosa* aux carbapénèmes est principalement due à des mutations chromosomiques, menant à la perte ou l'inactivation de la porine OprD et/ou à l'hyperexpression des pompes d'efflux [115]. Les principaux facteurs de risque liés à cette résistance dans la littérature sont l'exposition aux antibiotiques (surtout aux carbapénèmes), la présence de matériel, la présence de comorbidités et le passage en soins intensifs [116]. Aussi, même si les caractéristiques de la population couverte par les données de l'ECDC pour la France ne sont pas disponibles, nos résultats reflètent probablement une incidence plus élevée de bactériémies nosocomiales, d'exposition aux soins et aux antibiotiques dans la population de notre base. En effet, les 14 hôpitaux inclus sont affiliés à des universités, et plusieurs sont des centres d'expertise en maladies infectieuses. Ainsi, ils traitent une population de patients particulière et probablement plus comorbide que les autres centres de santé de la région. Malgré ces différences, l'évolution temporelle de l'antibiorésistance au cours de la période de l'étude était cohérente entre la base de données BactHub, les données de la France pour l'ECDC, et les données globales de l'ECDC (**Annexe 10**) [18,37].

Alors que la plupart des bases de données sur les bactériémies sont purement microbiologiques, la base Bacthub contient des données cliniques, ce qui nous a permis d'estimer une incidence au niveau multi-hospitalier (20-22 séjours avec bactériémie / 1 000 séjours). Cette incidence était plus élevée que celles d'anciens travaux multicentriques [89,117], ce qui pourrait s'expliquer par l'augmentation progressive de l'incidence des bactériémies au fil des années, en lien avec l'augmentation de l'espérance de vie, les

personnes âgées étant les plus à risque [78,117]. Néanmoins, ces incidences pourraient aussi s'expliquer par les particularités de la population traitée dans les hôpitaux inclus.

Peu d'études ont utilisé des données d'EDS pour étudier les infections aiguës, notamment les infections virales (SARS-CoV-2 [118], VIH [119], VHC [75]) ou les bactériémies [71,72]. Les données hospitalières présentent généralement de nombreux problèmes de qualité liés à l'hétérogénéité, le manque d'exhaustivité, et le manque de structuration des données, et à la multiplicité des logiciels utilisés en pratique clinique [76], affectant la validité des études. Dans notre cas, le fait de travailler sur les bactériémies a facilité l'étape du diagnostic puisqu'il se base sur des données microbiologiques. De plus, nous avons contrôlé le biais de recrutement en incluant uniquement les hôpitaux avec une bonne remontée des données de microbiologie, et une stabilité de cette remontée dans le temps. Après avoir structuré la base de données, nous avons décrit les caractéristiques principales des bactériémies, notamment les caractéristiques des patients et des épisodes, des isolats bactériens et leur résistance aux antibiotiques. La concordance de nos principaux résultats avec la littérature nous rassure quant à son utilisation à des fins de recherche.

La base présente plusieurs limites. En effet, en lien avec des données inégalement disponibles entre les hôpitaux, le score de sévérité de SOFA n'a pu être calculé, et le traitement des épisodes n'a pu être décrit. Par ailleurs, notre définition des infections communautaires excluait les patients sortis de l'un des hôpitaux de l'AP-HP depuis moins de 7 jours. Comme les hôpitaux de l'AP-HP ne couvrent qu'un tiers des séjours de soins aigus en Ile-de-France [106], il est possible que certains patients aient été classés à tort comme ayant une infection communautaire. En outre, nous n'avons pas été en mesure de différencier, au sein des infections communautaires, celles associées aux soins de celles communautaires pures. Ainsi, bien que proches des taux rapportés dans la mission PRIMO, nos taux de SARM semblaient élevés au regard de la faible circulation de souches épidémiques de SARM dans la communauté en France [120]. Enfin, nous avons utilisé les codes CIM-10 pour évaluer les sites primaires des bactériémies [29,97,98]. Avec cette méthode, un quart des épisodes n'avaient pas de site primaire et ont donc été classés comme des bactériémies primaires. Ce taux est supérieur à celui rapporté par deux précédents travaux réalisés sur moins de 4000 patients mais s'appuyant sur des critères cliniques et microbiologiques pour déterminer le site primaire [89,90]. Dans notre base, le taux de sites primaires multiples était très élevé (24-35%) comparé aux 4% de l'étude de Brun-Buisson et al [89]. En accord avec les comparaisons réalisées en partie 5.2.b, notre taux est certainement surestimé. Néanmoins, la répartition des sites primaires était comparable entre les CIM-10 et la lecture experte, après exclusion des bactériémies primaires et secondaires avec site multiple, ce qui conforte nos résultats. De plus, la distribution des sites d'infection primaire était cohérente avec la distribution des isolats bactériens.

Au total, malgré ces limites, la base BactHub contient des données pour près de 30 000 patients adultes hospitalisés avec une bactériémie entre 2016 et 2019 dans 14 hôpitaux universitaires avec une activité de soins aigus de la région du Grand Paris. Cette base est la première en France à présenter des données cliniques et microbiologiques détaillées de bactériémies au sein d'un groupe multi-hospitalier, et l'une des rares à l'échelle internationale [121–123]. L'analyse des principales caractéristiques des patients, épisodes, et isolats bactériens nous a permis de valider les données de la base et conforte son utilisation pour la recherche.

7 RÉSISTANCE AUX ANTIBIOTIQUES ET RÉCURRENCE DE BACTÉRIÉMIE A 1 AN DANS LES BACTÉRIÉMIES COMMUNAUTAIRES

7.1 Introduction

La récurrence de bactériémie est un évènement significatif qui s'accompagne d'une morbi-mortalité importante et entraîne une nouvelle exposition à des antibiotiques. Elle désigne la survenue d'un nouvel épisode de bactériémie après un premier épisode. Ce nouvel épisode infectieux peut être soit une rechute de la même infection qui n'était pas guérie, soit une nouvelle infection, à la même bactérie ou à une autre bactérie. D'après la littérature, la définition de la récurrence ne fait pas consensus, en effet, un épisode de bactériémie y est défini comme récurrent si il survient de 2 à 30 jours après un premier épisode [45,72,110,111,124–139]. De même, la période prospective pendant laquelle est étudiée la récurrence s'échelonne entre 30 jours et 14 ans après le premier épisode [45,72,110,111,124–139]. En conséquence, les taux de récurrence de bactériémie décrits varient de 3,7 à 37% [45,72,110,111,124–137].

Si l'on se restreint à la récurrence à un an, et aux études ne ciblant pas une population particulière (avec comorbidités par exemple) ou une bactériémie à une bactérie donnée, le taux de récurrence variait entre 9,2% et 12% [110,111,126]. Plusieurs facteurs de risque associés à la récurrence de bactériémie à un an ont été identifiés dans la littérature. Ces facteurs étaient liés aux comorbidités des patients (score de Charlson > 0 [111,126]), aux caractéristiques du premier épisode (caractère associé aux soins [111,126], absence de site primaire ou sites primaires multiples [126], site primaire digestif ou lié à un cathéter central, ou ostéoarticulaire, ou une endocardite [126]), aux bactéries identifiées lors du premier épisode (épisode plurimicrobien ou à champignon [111,126]), et au traitement du premier épisode (traitement empirique²⁸ inadapté [110,126]).

Les infections bactériennes résistantes aux antibiotiques pourraient être associées à un risque de récurrence plus élevé. En effet, l'antibiorésistance pourrait s'associer à des infections plus graves, un allongement des durées de séjour, et augmenter la susceptibilité des hôtes à l'infection [28,31]. Cinq études ont évalué l'antibiorésistance en tant que potentiel facteur de risque de récurrence de bactériémie [132–134,139,140]. Toutes ont ciblé des espèces particulières et n'ont étudié que la récurrence à cette espèce. Des associations brutes entre certains couples bactérie-résistance (dont les entérobactéries

²⁸ Traitement débuté avant d'avoir une information sur le pathogène

résistantes aux C3G et le SARM), et la récurrence d'une bactériémie à la même espèce ont été identifiées [132,133]. Néanmoins, après ajustement, l'étude de Choi et al., centrée sur le SARM, conclut à une association non significative [133]. Par ailleurs, aucune association entre la résistance et la récurrence n'a été retrouvée pour les bactériémies à pneumocoque [134], à *Enterobacter spp.* [140], et à *E. coli* ou *K. pneumoniae* BLSE [139].

Enfin, aucune étude ne s'est intéressée spécifiquement aux facteurs de risque de récurrence après une bactériémie communautaire. Ce point semble d'intérêt, car les populations présentant des bactériémies communautaires et nosocomiales sont différentes en termes d'âge, de comorbidités, et d'épidémiologie bactérienne (cf. partie 6 du manuscrit). Par ailleurs, la diffusion croissante des BLSE dans la communauté incite à mieux cibler les populations à risque de récurrence en communautaire.

L'étude présentée ici avait pour objectif d'évaluer l'impact de l'antibiorésistance sur le risque de récurrence de bactériémie à 12 mois après une bactériémie communautaire, en prenant en compte d'autres facteurs de risque ou de confusion potentiels. Nous nous sommes réduits aux espèces bactériennes *S. aureus*, *E. coli* et *Klebsiella spp.*, afin d'étudier l'impact de résistances significatives comme la résistance à la méticilline de *S. aureus*, et la résistance aux C3G d'*E. coli* et *Klebsiella spp.*, dont le mécanisme majoritaire est la production de BLSE [55,141].

7.2 Matériels et Méthodes

Population d'étude

Pour chaque patient, le premier épisode communautaire monomicrobien à *S. aureus*, *E. coli* ou *Klebsiella spp.* entre le 1^{er} janvier 2017 et le 31 décembre 2019 dans l'un des 14 hôpitaux de l'AP-HP ont été identifiés. Afin de ne considérer que les épisodes incidents, les épisodes avec un antécédent de bactériémie (communautaire ou nosocomiale) dans les 12 mois ont été exclus, quel que soit le germe. Enfin, seuls les épisodes où le patient a survécu ont été retenus. La récurrence était définie par la survenue d'un épisode de bactériémie entre 7 et 365 jours après l'épisode incident. Les épisodes survenus dans les 7 jours étaient exclus pour éviter d'inclure des rechutes précoces ; en effet les recommandations des durées de traitement des infections communautaires dépassent rarement deux semaines dans les recommandations françaises [142]. Les épisodes récurrents pouvaient être communautaires ou nosocomiaux, et à n'importe quelle espèce bactérienne. Par la suite deux groupes ont été considérés : avec ou sans récurrence.

Données recueillies

Pour chaque épisode incident, le sexe, l'âge, les dates d'admission et de sortie, la date de décès (le cas échéant), la durée de l'épisode (de la première hémoculture positive à la fin du séjour), le parcours de soin hospitalier, et les résultats de microbiologie (bactéries identifiées et résultats de sensibilité aux antibiotiques) ont été recueillis.

Pour chaque épisode récurrent, les dates d'admission et de sortie, le caractère communautaire ou nosocomial, et les résultats de microbiologie ont été collectés.

Variables étudiées

Les variables étudiées pouvaient être liées aux patients ou à leur épisode incident. Les variables liées aux patients étaient le sexe (masculin / féminin), l'âge (considéré en 5 classes : 18-35 ans, 35-50 ans, 50-65 ans, 65-80 ans, > 80 ans), le score de Charlson (considéré en 4 classes : 0, 1-2, 3-4, >4) et les comorbidités (cancer, insuffisance cardiaque, diabète, maladies vasculaires, rénales, hépatiques, systémiques, maladies pulmonaires chroniques, démence, hémiplégie). Pour chaque épisode incident, la durée a été considérée en 4 classes : ≤ 7 jours, 7-14 jours, 14-30 jours, 30-60 jours. Le passage en soins intensifs, le passage en chirurgie, et la présence d'un choc septique (code CIM-10, **Annexe 4**) ont aussi été retenus. Le traitement empirique a été étudié sur une fenêtre de 24h encadrant la date de prélèvement de la première hémoculture positive de l'épisode. Pour chaque épisode, le traitement empirique était considéré comme adapté ou inadapté, en fonction des résultats de sensibilité aux antibiotiques de la bactérie isolée, suivant les recommandations de la CA-SFM [96]. Le site primaire de la bactériémie a été considéré en 8 classes : absence, urinaire, respiratoire bas, digestif, cardio-vasculaire, sur matériel, autres, et sites primaires multiples. Les épisodes incidents ont été classifiés en primaires ou secondaires selon l'absence ou la présence d'un site primaire, respectivement. La bactérie et la résistance pouvant être liées, une variable composite en six classes a été créée : *S. aureus* sensible à la méticilline (SAMS), SARM, *E. coli* non C3G-R, *E. coli* C3G-R, *Klebsiella spp* non C3G-R, *Klebsiella spp* C3G-R.

Analyses statistiques

Les patients, les épisodes incidents, les bactéries isolées lors des épisodes incidents et leur antibiorésistance, ont été décrits en fonction de la présence ou non de récurrence. Une brève description des premiers épisodes récurrents et des bactéries isolées lors de ces épisodes est aussi présentée.

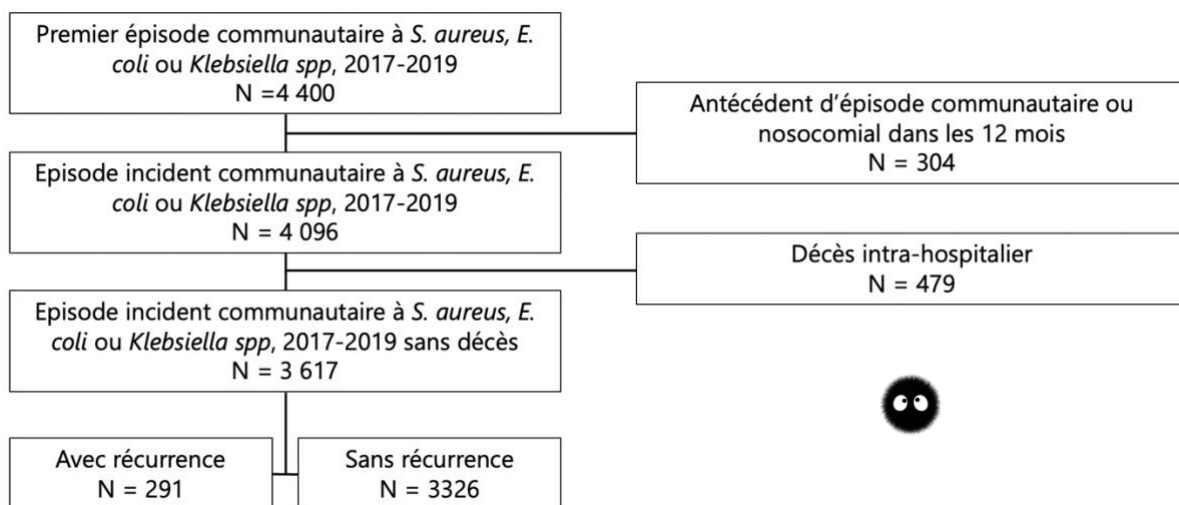
Des modèles de régression logistique univariée et multivariée ont été réalisés pour identifier les facteurs de risque de récurrence après un épisode incident. Les variables dont la p-value était $\leq 0,20$ dans les analyses univariées étaient incluses dans le modèle multivarié. Les variables retenues dans le modèle final ont été sélectionnées par une méthode pas à pas descendante, en forçant les variables âge et sexe qui

sont habituellement associées aux infections bactériennes. Une analyse de sensibilité prenant en compte l'adéquation du traitement empirique, sur l'échantillon d'épisodes où cette information est disponible, a été réalisée, avec les mêmes variables que celles retenues dans le modèle final. HiveQL, Python 3, PySpark 2.4.3, et R 4.0.0 ont été utilisés pour effectuer les analyses statistiques.

7.3 Résultats

Entre 2017 et 2019, 4 400 patients avaient au moins un épisode communautaire à *S. aureus*, *E. coli* ou *Klebsiella spp.* (**Figure 7.1**). Seuls les premiers épisodes ont été retenus et les épisodes avec un antécédent de bactériémie dans les 12 mois (n = 304) ont été exclus. Parmi ces épisodes, 11,7% (n=479) se sont terminés par un décès ; la mortalité intra-hospitalière était similaire à celle rapportée sur la totalité des premiers épisodes de la base dans la partie 6.2. Au total, 3 617 épisodes incidents ont été retenus, parmi lesquels 8% (n = 291) ont été suivis d'au moins un épisode récurrent.

Figure 7.1 Identification des épisodes incidents communautaires à *S. aureus*, *E. coli* ou *Klebsiella spp.*, BactHub 2017-2019, sans décès



Analyses descriptives

Épisodes incidents

Les patients avec récurrence étaient plus souvent de sexe masculin (56,7% vs 47,9%, p = 0,004). La répartition des âges différait entre les deux groupes (p = 0,0005), avec plus de patients entre 50-80 ans (66,7% vs 55,2%) et moins de patients >80 ans (18,2% vs 26,6%) dans le groupe récurrent (**Tableau 7.1**). Le score de Charlson différait entre les deux groupes (p < 0,0001) ; il était nul pour 44% des patients sans

réurrence, et 27% des patients avec réurrence. Les patients avec réurrence avaient deux fois plus de cancer (37,6% vs 19,5%, $p < 0,0001$), et présentaient plus de maladies rénales (22,3% vs 13,6%, $p = 0,002$) et hépatiques (13,8% vs 7,5%, $p = 0,0006$), que les patients sans réurrence. La proportion des autres comorbidités ne différait pas significativement entre les deux groupes. La durée médiane de l'épisode incident était de 8,0 [4,0 – 15,5] jours dans le groupe récurrent, et de 7,0 [3,0 – 15,0] jours dans le groupe sans réurrence ; la distribution de la durée était similaire ($p = 0,29$). Le taux de passage en chirurgie, en soins intensifs, et de survenue d'un choc septique ne différaient pas entre les deux groupes. La répartition des sites primaires différait significativement entre les deux groupes ($p < 0,0001$). Les bactériémies primaires (24,1% vs 14,0%) et les bactériémies secondaires associées à une infection digestive (12,1% vs 9,5%) ou sur matériel (7,8% vs 4,4%) semblaient plus fréquentes dans le groupe avec réurrence. A contrario, les sites primaires urinaires semblaient plus fréquents dans le groupe sans réurrence (35,7% vs 26,2%). La distribution des bactéries et des résistances différait entre les deux groupes ($p < 0,0001$). Il semblait y avoir plus de *Klebsiella spp.* C3G-R (7,2% vs 1,9%) ou non C3G-R (13,1% vs 9,3%) et plus d'*E. coli* C3G-R (13,1% vs 7,6%) dans le groupe récurrent. Près de la moitié des isolats étaient des *E. coli* non C3G-R dans les deux groupes.

Épisodes récurrents

Les patients avec réurrence ont présenté en moyenne 2,25 épisodes (min – max : 2 – 8) sur la période de l'étude, incluant leur épisode incident. Les premiers épisodes récurrents survenaient dans un délai médian de 80 jours [Q1-Q3 : 30,0 – 175,0] après l'épisode incident, et étaient majoritairement communautaires ($n = 166/291$, 57,0%). La durée médiane des premiers épisodes récurrents était de 11 jours [Q1-Q3 : 6,0 – 19,0]. Elle était supérieure à la durée médiane des épisodes incidents, probablement du fait de la présence d'épisodes récurrents nosocomiaux. Les épisodes récurrents étaient à la même espèce bactérienne que les épisodes incidents dans 47,4% des cas ($n = 138$). Ce taux était plus élevé lorsque l'épisode incident était à *E. coli* ($n = 92/174$; 52,9%) que lorsqu'il était à *Klebsiella spp.* ($n = 25/59$; 42,4%) ou à *S. aureus* ($n = 21/58$; 36,2%).

Tableau 7.1 Caractéristiques des patients et des épisodes incidents communautaire à *S. aureus*, *E. coli* ou *Klebsiella spp*, avec et sans récurrence, base BactHub, 2017-2019

	Avec récurrence, N = 291	Sans récurrence, N = 3 326	p-value
Patients			
Sexe, N (%)			0,004
Homme	165 (56,7)	1594 (47,9)	
Femme	126 (43,3)	1732 (52,1)	
Age, années			0,0005
18-35	10 (3,4)	227 (6,8)	
35-50	34 (11,7)	378 (11,4)	
50-65	84 (28,9)	765 (23,0)	
65-80	110 (37,8)	1070 (32,2)	
>80	53 (18,2)	886 (26,6)	
Score de Charlson*			<0,0001
0	75 (26,6)	1389 (44,1)	
1-2	109 (38,7)	1004 (31,9)	
>2	98 (34,7)	759 (24,0)	
Comorbidités*, N (%)			
Cancer	106 (37,6)	615 (19,5)	<0,0001
Insuffisance cardiaque	35 (12,4)	423 (13,4)	0,63
Diabète	63 (22,3)	730 (23,2)	0,75
Maladie vasculaire	23 (8,2)	340 (10,8)	0,15
Maladie rénale	63 (22,3)	430 (13,6)	0,002
Maladie hépatique	39 (13,8)	237 (7,5)	0,0006
Maladie pulmonaire chronique	12 (4,3)	189 (6,0)	0,21
Démence	9 (3,2)	178 (5,7)	0,06
Paralysie (hémiplégie / paraplégie)	4 (1,4)	84 (2,7)	0,17
Maladie systémique	5 (1,8)	32 (1,0)	0,28
Épisodes incidents			
Durée (jours)			0,29
≤7	139 (47,8)	1730 (52,0)	
7-14	74 (25,4)	706 (21,2)	
14-30	50 (17,2)	613 (18,4)	
30-60	28 (9)	277 (8,3)	
Passage en chirurgie**, N (%)	37 (12,9)	426 (12,9)	1,00
Passage en soins intensifs**, N (%)	70 (24,4)	718 (21,7)	0,30
Choc septique*, N (%)	27 (9,6)	279 (8,9)	0,69
Site d'infection primaire*, N (%)			<0,0001
Bactériémie primaire [#]	68 (24,1)	442 (14,0)	
Bactériémie secondaire			
Infections multiples	59 (20,9)	747 (23,7)	
Infection urinaire	74 (26,2)	1125 (35,7)	
Infection respiratoire basse	14 (5)	190 (6,0)	

Infection digestive	34 (12,1)	300 (9,5)
Infection sur matériel	22 (7,8)	137 (4,4)
Autres	11 (3,9)	211 (6,7)
<hr/>		
Bactérie et résistance, N (%)		<0,0001
SAMS	54 (18,6)	737 (22,2)
SARM	4 (1,4)	75 (2,3)
<i>E. coli</i> non C3G-R	136 (46,7)	1889 (56,8)
<i>E. coli</i> C3G-R	38 (13,1)	253 (7,6)
<i>Klebsiella spp</i> non C3G-R	38 (13,1)	310 (9,3)
<i>Klebsiella spp</i> C3G-R	21 (7,2)	62 (1,9)

Abréviations : C3G-R : résistant aux céphalosporines de 3^{ème} génération ; Q1 : premier quartile ; Q3 : troisième quartile ; SAMS / SARM : *S. aureus* sensible / résistant à la méticilline. * Données manquantes : 9 épisodes avec récurrence, 174 épisodes sans récurrence. ** Données manquantes : 4 épisodes avec récurrence, 19 épisodes sans récurrence. # Absence de site primaire.

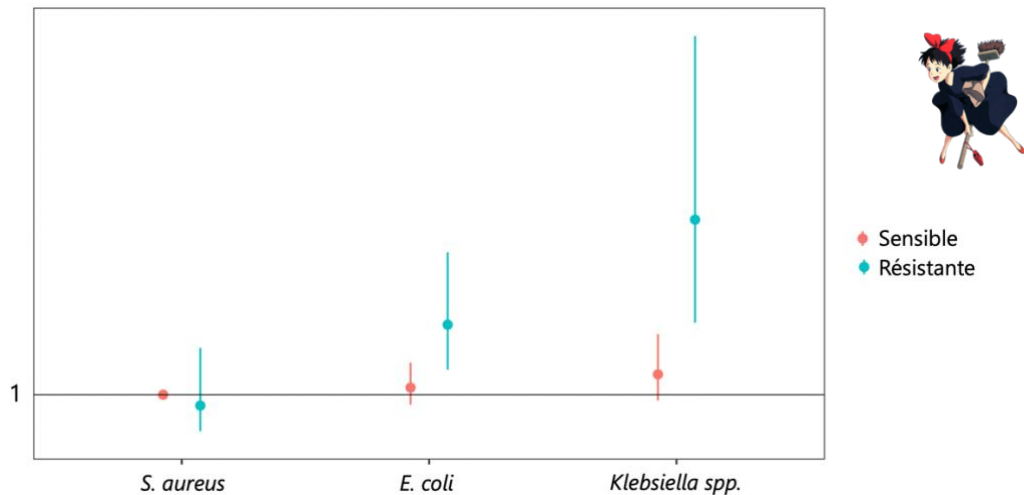
Modèles logistiques

Les résultats des analyses univariées et multivariées pour identifier les facteurs associés à la récurrence à 1 an sont présentés dans le **Tableau 7.2**. Les variables non considérées dans l'analyse multivariée étaient : une comorbidité de type insuffisance cardiaque, diabète, maladie systémique, la durée de l'épisode incident, le passage en chirurgie, le passage en soins intensifs, la présence d'un choc septique. L'inclusion des comorbidités a été préférée à celle du score de Charlson. Les variables non liées à la récurrence, et donc non retenues dans le modèle final, étaient les maladies vasculaires, les maladies pulmonaires chroniques, la démence, et la présence d'une paralysie.

Les résultats montrent que comparé au SAMS, l'identification d'un *E. coli* non C3G-R ou d'une *Klebsiella spp.* non C3G-R, n'est pas liée à la récurrence. Pour *S. aureus*, la présence d'une résistance à la méticilline n'est pas liée à la récurrence. A contrario, la résistance aux C3G, pour les entérobactéries étudiées, joue un rôle important dans le risque de récurrence. Ainsi, l'odds ratio de *Klebsiella spp.* C3G-R (OR 4,7 [2,5 – 8,6]) est trois fois plus élevé que celui de *Klebsiella spp.* non C3G-R (OR 1,4 [0,9 - 2,3], **Figure 7.2**). De même, l'odds ratio d'*E. coli* C3G-R (OR 2,5 [1,5 – 4,0]) est deux fois plus élevé que celui d'*E. coli* non C3G-R (1,2 [0,8 - 1,7]). Ces résultats sont confirmés quel que soit la référence choisie (**Annexe 11**) et par une analyse stratifiée sur le germe (non présentée). Une analyse de sensibilité sur les épisodes avec une information sur le traitement empirique (n = 1292 ; *S. aureus* n = 248 SA, *E. coli* n = 892, *Klebsiella spp.* n = 152), montre que l'adéquation du traitement empirique n'a pas d'effet sur le risque de récurrence. Par ailleurs, les conclusions sur l'impact de la résistance sur la récurrence sont similaires (**Annexe 12**), mais montrent un OR plus important associé à *Klebsiella spp.* C3G-R, probablement du fait de faibles effectifs. Certains sites primaires ont été identifiés comme étant des facteurs de risque indépendants de récurrence : l'absence de site primaire (OR 2,4 [1,7 – 3,5]), les infections sur matériel (OR 2,0 [1,1 – 3,4]), ou digestives (OR 1,6 [1,0 – 2,5]). Enfin, certaines comorbidités étaient également liées à la récurrence :

les cancers (OR 2,2 [1,7 – 2,9]), les maladies rénales (OR 1,8 [1,3 – 2,5]) et les maladies hépatiques (OR 1,8 [1,2 – 2,6]).

Figure 7.2 Odds ratio ajustés et intervalles de confiance à 95% du lien entre chaque couple bactérie-résistance et la récurrence de bactériémie à 1 an



La catégorie sensible correspondait à une sensibilité à la méticilline pour *S. aureus*, et une sensibilité aux C3G pour *E. coli* et *Klebsiella spp.* La catégorie résistante correspondait à une résistance à la méticilline pour *S. aureus*, et une résistance aux C3G pour *E. coli* et *Klebsiella spp.*

Tableau 7.2 Analyses univariées et multivariées des facteurs de risque de récurrence à 1 an suite à un épisode incident communautaire monomicrobien à *S. aureus*, *E. coli* ou *Klebsiella spp.*

	Analyse univariée		Analyse multivariée	
	OR [IC95]	p-value	OR [IC95]	p-value
Patients				
Sexe (ref = homme)	0,70 [0,55 – 0,89]	0,004	0,93 [0,71 – 1,21]	0,56
Age (ref = 35-50)		0,0005		0,06
18-35	0,49 [0,23 – 0,97]		0,62 [0,27 – 1,30]	
50-65	1,22 [0,81 – 1,87]		1,14 [0,74 – 1,80]	
65-80	1,14 [0,77 – 1,73]		1,12 [0,73 – 1,75]	
>80	0,67 [0,43 – 1,05]		0,76 [0,47 – 1,23]	
Score de Charlson (ref = 0)		<0,0001		
1-2	2,01 [1,49 – 2,74]			
>2	2,39 [1,75 – 3,28]			
Comorbidités				
Cancer	2,48 [1,92 – 3,21]	<0,0001	2,19 [1,66 – 2,88]	<0,0001
Insuffisance cardiaque	0,91 [0,62 – 1,30]	0,63		
Diabète	0,95 [0,71 – 1,27]	0,75		
Maladie vasculaire	0,73 [0,46 – 1,12]	0,15		
Maladie rénale	1,64 [1,20 – 2,21]	0,002	1,78 [1,28 – 2,45]	0,0005
Maladie hépatique	1,97 [1,36 – 2,81]	0,0006	1,80 [1,21 – 2,62]	0,006
Maladie pulmonaire chronique	0,70 [0,36 – 1,21]	0,21		

Démence	0,55 [0,26 – 1,03]	0,06	
Paralysie (hémiplégie, paraplégie)	0,53 [0,16 – 1,27]	0,17	
Maladie systémique	1,76 [0,60 – 4,17]	0,28	
Épisodes incidents			
Durée (ref = 7-14)		0,29	
≤7	0,77 [0,57 – 1,03]		
14-30	0,78 [0,53 – 1,13]		
30-60	0,96 [0,60 – 1,51]		
Passage en chirurgie	1,00 [0,69 – 1,42]	1,00	
Passage en soins intensifs	1,16 [0,87 – 1,53]	0,30	
Choc septique	1,09 [0,70 – 1,62]	0,69	
Site d'infection primaire (ref = Infection urinaire)		<0,0001	0,0003
Bactériémie primaire	2,34 [1,65 – 3,31]		2,41 [1,66 – 3,49]
Bactériémie secondaire			
Infections multiples	1,20 [0,84 – 1,71]		1,24 [0,84 – 1,81]
Infection respiratoire basse	1,12 [0,6 – 1,96]		1,25 [0,65 – 2,24]
Infection digestive	1,72 [1,11 – 2,62]		1,61 [1,02 – 2,49]
Infection sur matériel	2,44 [1,44 – 4,00]		1,97 [1,10 – 3,41]
Autres	0,79 [0,39 – 1,46]		0,98 [0,47 – 1,89]
Bactérie et résistance (ref = SAMS)		<0,0001	<0,0001
SARM	0,73 [0,22 – 1,84]		0,77 [0,23 – 1,99]
<i>E. coli</i> non C3G-R	0,98 [0,71 – 1,37]		1,15 [0,79 – 1,68]
<i>E. coli</i> C3G-R	2,05 [1,31 – 3,17]		2,48 [1,53 – 4,01]
<i>Klebsiella spp</i> non C3G-R	1,67 [1,08 – 2,58]		1,43 [0,88 – 2,28]
<i>Klebsiella spp</i> C3G-R	4,62 [2,58 – 8,06]		4,70 [2,52 – 8,58]

Abréviations : C3G-R : résistant aux céphalosporines de 3^{ème} génération ; SAMS / SARM : *S. aureus* sensible / résistant à la méticilline.

7.4 Discussion

Dans cette étude, nous montrons pour la première fois que la résistance aux antibiotiques dans les bactériémies communautaires est un facteur important de récurrence de bactériémie à 1 an. Par ailleurs, l'impact de cette résistance est fonction du germe. L'isolement d'une *Klebsiella spp.* C3G-R était associé au plus grand risque de récurrence, suivi de l'isolement d'un *E. coli* C3G-R, alors que l'identification d'un SARM n'était pas liée à la récurrence. Par ailleurs, nos résultats confortent dans la sous-population des bactériémies communautaires le lien entre les comorbidités des patients, le site primaire de la bactériémie, et le risque de récurrence. Nous montrons un impact différentiel des comorbidités, la présence d'un cancer étant la comorbidité la plus fortement liée à la récurrence, suivie par les maladies hépatiques ou rénales. Parmi les sites primaires, l'absence de site primaire, un site primaire digestif ou sur matériel étaient liés à la récurrence.

Peu de travaux ont étudié le lien entre l'antibiorésistance et la récurrence dans les bactériémies, et aucun ne s'est focalisé sur le contexte communautaire. Woudt et al. ont montré un lien entre l'isolement d'un SARM ou d'une entérobactérie résistante aux C3G, et la récurrence à la même espèce, avec des risques relatifs bruts inférieurs à 2 [132]. Notre étude montre que l'impact de la résistance est majeur après prise en compte des autres facteurs de risque, et identifie un effet différentiel du germe. La présence d'une résistance aux C3G multipliait par trois l'OR lié à la récurrence pour *Klebsiella spp.* et par deux l'OR pour *E. coli*. De plus, le lien entre la résistance aux C3G et la récurrence était 1,5 fois plus fort pour *Klebsiella spp.* que pour *E. coli*. En accord avec Choi et al., nous ne retrouvons pas d'effet du SARM sur la récurrence, malgré un suivi plus court et le caractère communautaire du séjour incident [133]. Nos résultats confirment la morbidité des infections à *E. coli* et surtout à *Klebsiella spp.* résistante aux C3G, dont la proportion est croissante dans le milieu communautaire [56,143]. Sans s'intéresser à la résistance, Al-Hasan et al. ont montré que l'isolement d'une *Klebsiella spp.* lors d'un épisode de bactériémie était associé à la récurrence, en comparaison avec l'isolement d'un *E. coli*, et après ajustement [111,124,126]. D'autres travaux suggèrent que les patients avec une bactériémie à *K. pneumoniae* BLSE présentent plus de passages en soins intensifs et une mortalité accrue, en comparaison avec les bactériémies à *E. coli* BLSE [144,145]. Nos résultats apportent de nouvelles données incitant à une vigilance accrue vis-à-vis de la diffusion des *Klebsiella spp.* résistantes aux antibiotiques en communautaire, et à différencier les espèces d'entérobactéries dans de futures études.

Cette étude confirme le lien entre la présence de comorbidités et la récurrence. Les odds ratio bruts liés au score de Charlson étaient similaires à ceux décrits par Jensen et al [126]. Bien que le score de Charlson ait été associé au risque de récurrence dans deux études [111,126], il n'a pas été considéré dans notre modèle final. En effet, il semblait intéressant d'identifier les comorbidités spécifiques liées à la récurrence, afin d'orienter les modalités de suivi des patients. De telles données sont à ce jour absentes pour les bactériémies communautaires, et les études s'intéressant aux comorbidités individuellement conduisaient à des résultats variables [72,129,133,134,136,139,140]. Par ailleurs, en accord avec Jensen et al., nous avons montré que l'absence de site primaire, et de manière secondaire la présence d'un site primaire digestif ou sur matériel, étaient liés à la récurrence [126]. Enfin, le taux de récurrence de bactériémie à 1 an plus faible dans notre étude (8%) comparé à d'autres travaux de la littérature (9,2–12%), pourrait s'expliquer par la sélection d'épisodes incidents communautaires [110,111,126].

Les forces de notre étude incluaient la taille de notre cohorte, ainsi que la richesse et le détail des informations cliniques et microbiologiques disponibles, permettant d'évaluer un nombre important de potentiels facteurs de risque. Malgré une cohorte de taille importante, les effectifs de bactéries résistantes

dans le groupe avec récurrence sont faibles. Néanmoins ces effectifs conduisent à des résultats significatifs et montrent un impact de la résistance, sur la récurrence, qui dépend du germe. Le choix du suivi sur 1 an, sans restriction sur l'espèce identifiée lors de la récurrence, tient compte du déséquilibre prolongé que peut induire l'exposition aux antibiotiques sur la flore microbienne, et des durées de portage chronique des bactéries résistantes aux antibiotiques, pouvant atteindre un an pour les BLSE [146–150].

En conclusion, dans cette étude multicentrique s'appuyant sur la base BactHub, nous avons montré sur un échantillon de plus de 3 500 épisodes incidents de bactériémie communautaire que la résistance aux C3G de *Klebsiella spp.* ou *E. coli* augmentait de manière non négligeable le risque de récurrence. Ce risque était maximal pour *Klebsiella spp.* C3G-R, dont la diffusion en communautaire représente un réel problème de santé publique. Cette étude illustre une autre facette de la morbi-mortalité induite par les entérobactéries résistantes aux antibiotiques.

8 EXPOSITION INDIVIDUELLE AUX ANTIBIOTIQUES ET INFECTIONS BACTÉRIENNES RÉSISTANTES AUX ANTIBIOTIQUES

8.1 Introduction

Le lien entre l'exposition aux antibiotiques et l'antibiorésistance a été analysé par de nombreuses études épidémiologiques, principalement de type cas-témoins. Ces études présentent une hétérogénéité importante dans la méthode, en particulier dans les définitions du groupe témoin et de l'exposition aux antibiotiques [151–155]. Par ailleurs, la plupart des études s'intéressaient aux infections nosocomiales, et beaucoup ne différenciaient pas colonisation et infection [12]. Dans une étude, Olesen et al. ont montré qu'à l'échelle populationnelle, le développement de la résistance serait d'avantage lié à l'utilisation large des antibiotiques qu'à l'intensité de leur utilisation [156]. La majorité des antibiotiques étant dispensée en ville, et la résistance se développant en communauté, des études permettant de caractériser plus finement le lien entre exposition aux antibiotiques et antibiorésistance dans les infections communautaires sont nécessaires. Par ailleurs, ce lien doit être évalué en fonction du couple bactérie-résistance. En effet, pour une même souche bactérienne, différents mécanismes interagissent au niveau individuel et populationnel, conduisant à l'émergence d'une résistance particulière chez un hôte donné [157]. De plus, la littérature tend à montrer que les résultats observés pour une espèce bactérienne ne peuvent être extrapolés à une autre, voire à différentes souches au sein d'une même espèce bactérienne [157].

Dans cette partie, nous passons en revue les études cas-témoins évaluant l'impact de l'exposition aux antibiotiques sur la survenue d'infections urinaires (IU) ou de bactériémies communautaires résistantes aux antibiotiques. Dans les études sélectionnées, nous montrons l'hétérogénéité dans les points clés liés au facteur de risque d'intérêt, à la population et au schéma de l'étude : exposition aux antibiotiques, cas, témoins et schéma cas-témoins. Nous discutons de l'impact de cette hétérogénéité sur les résultats des études, en nous concentrant sur les E-BLSE, souvent étudiées en raison de leur propagation rapide dans la communauté, de leur propension à coloniser le microbiote humain et de leur morbidité et mortalité [158]. Par la suite, nous proposons de définir une hypothèse physiopathologique spécifique au couple bactérie-résistance étudié afin de clarifier la conception et l'interprétation des études futures. Enfin, nous discutons l'apport potentiel de la base BactHub pour l'étude du lien entre l'exposition aux antibiotiques et le développement d'infections communautaires résistantes aux antibiotiques.



8.2 Matériels et méthodes

Revue de la littérature

L'ensemble des études cas-témoins étudiant l'exposition aux antibiotiques comme facteur de risque d'antibiorésistance dans les bactériémies ou les infections urinaires (IU) communautaires hospitalisées ont été recherchées dans la littérature. Cette requête a été effectuée dans PubMed en juin 2020, sans restriction sur la langue ou la date de publication. J'ai examiné les titres et les résumés des articles sélectionnés et exclu ceux dont l'objectif n'était pas celui d'intérêt. Les études retenues étaient évaluées, et j'ai exclu (1) celles où les infections n'étaient pas communautaires ou n'étaient pas des IU ou des bactériémies ; (2) celles où le schéma d'étude n'était pas de type cas-témoin ; (3) les études sans analyse des facteurs de risque d'antibiorésistance, ou ne considérant pas l'exposition aux antibiotiques comme facteur de risque ; (4) les études sans analyse statistique, sans présentation des résultats, ou celles dont le texte intégral n'était pas accessible. J'ai ensuite analysé les articles retenus. Afin d'identifier d'autres articles non identifiés par ma requête, les références des articles sélectionnés ont été étudiées.

Facteur de risque : l'exposition aux antibiotiques

Quatre critères ont été considérés pour définir l'exposition aux antibiotiques [151] : la classification retenue pour les antibiotiques (globale, molécule, famille, ou spectre d'activité), (2) la source des données sur l'exposition (questionnaire, base médico-administrative, etc.), (3) la fenêtre d'exposition, c'est-à-dire la fenêtre précédant l'hospitalisation d'inclusion, au cours de laquelle l'exposition aux antibiotiques est évaluée (7 jours, 1, 3, 6, 12 mois, etc...), et (4) la mesure de l'indicateur qui peut être binaire (exposition / non exposition), à plusieurs classes, ou continue (nombre de doses définies journalières, durée du traitement, etc.).

Définitions et impacts sur les résultats

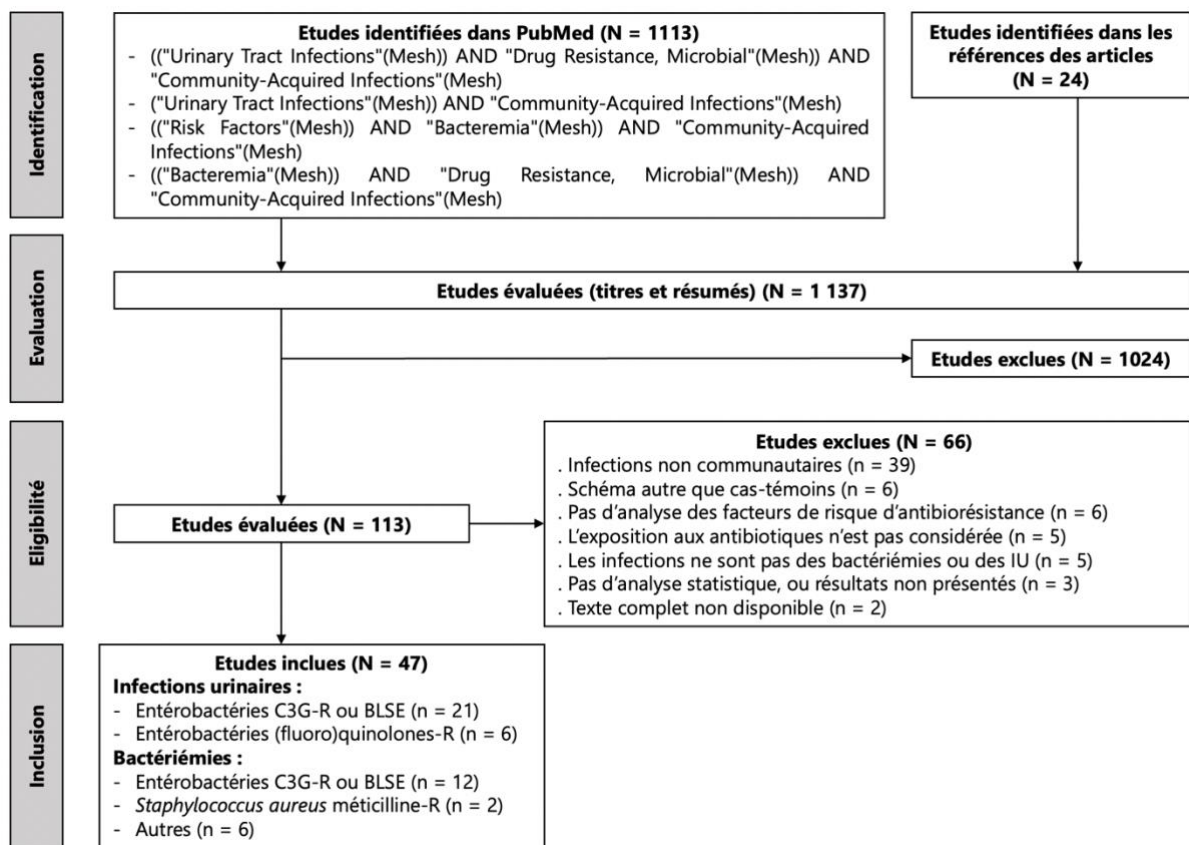
Dans chaque étude, le type d'infections communautaires considéré, et sa définition, ont été évalués : il pouvait s'agir des infections communautaires vraies, ou des infections communautaires. Une distinction a été établie entre le type d'infections communautaires considéré par l'étude, et celui qui correspond de manière effective à la définition utilisée. Par exemple, si une étude ciblant les infections communautaires vraies n'excluait pas les infections associées aux soins, nous y ferons référence dans la suite de ce manuscrit en indiquant « infections communautaires vraies (en réalité communautaires) ».

La définition des cas, des témoins et le schéma d'étude ont été recueillis. Les impacts de ces choix sur les résultats est discuté.

8.3 Résultats

Sur 1 137 articles identifiés, 113 étaient éligibles après examen de leurs titres et de leurs résumés. Parmi eux, 66 ont été exclus, principalement car ils ne concernaient pas les infections communautaires (39/66, 59%), et 47 ont été inclus à la revue de la littérature (bactériémies, N = 15 ; IU, N = 32 ; **Figure 8.1**) [159–205]. Le groupe d'intérêt dans chaque étude (infections communautaires ou communautaires vraies) et la paire bactérie-résistance étudiée sont présentés dans l'**Annexe 13**.

Figure 8.1 Diagramme PRISMA de sélection des études cas-témoins évaluant le lien entre l'exposition aux antibiotiques et la survenue d'une IU ou d'une bactériémie communautaire résistante aux antibiotiques



Abréviations : BLSE : β -lactamases à spectre étendu ; C3G-R : résistantes aux céphalosporines de 3^{ème} génération ; IU : infection urinaire ; R : résistant.

Exposition aux antibiotiques

Classification

Parmi les études retenues, la classification des antibiotiques était très hétérogène (**Tableau 8.1**). L'exposition globale aux antibiotiques était étudiée dans 77% (36/47) des études, tandis que les céphalosporines, les (fluoro)quinolones et les pénicillines étaient fréquemment individualisées, avec diverses sous-classifications. Certains antibiotiques étaient majoritairement étudiés dans les IU, car ils

font partie de l'arsenal thérapeutique de ces infections : cotrimoxazole, fosfomycine, nitrofurantoïne. Parmi toutes les études, 30% (14/47) ont individualisé au moins 5 différent(e)s (familles d') antibiotiques. Notons que, 9% (4/47) en ont individualisé plus de 10, sans mentionner de correction du seuil de significativité [174,184,195,201]. Les antibiotiques à spectre anti-anaérobies ont été individualisés dans 26 % (12/47) des études (« anti-anaérobies », N = 1 ; métronidazole, N = 2 ; pénicillines avec inhibiteurs de β -lactamase, N = 9).

Sources des données sur l'exposition

Les sources de données concernant l'exposition aux antibiotiques n'étaient précisées que dans 87% (41/47) des études (**Tableau 8.1**). Les dossiers médicaux étaient utilisés dans plus de la moitié des études (27/47). Les entretiens avec les patients et les bases de données des pharmacies servaient de base au recueil dans 13% (6/47) et 9 % (4/47) des études, respectivement.

Fenêtre d'exposition

Parmi les 47 études analysées, 2% (1/47) ne définissaient pas de fenêtre d'exposition, et les autres ont considéré une fenêtre d'exposition variant entre 15 jours et 1 an (**Tableau 8.1**). La moitié des études (24/47) limitaient la fenêtre d'exposition à 3 mois. Seules 21% des études (10/47) considéraient des fenêtres d'exposition supérieures à 3 mois. Une seule étude mentionnait explicitement l'exclusion de l'exposition récente aux antibiotiques, considérée dans le mois précédent l'inclusion des patients [176]. Une étude évaluait plusieurs fenêtres d'exposition [200].

Mesure

La majorité des études ont pris en compte l'exposition antérieure aux antibiotiques de façon binaire (exposition vs non-exposition). Parmi elles, deux études ont pris en compte dans la dichotomisation un ou plusieurs seuil(s) quantitatif(s) (exposition à au moins une dose standard pendant au moins 24 h [180], exposition à un antibiotique pendant ≥ 3 jours [171]). Une étude considérait l'exposition à travers des variables à plus de deux classes pour la dose et la durée du traitement [200].

Tableau 8.1 Classification, fenêtre d'exposition, et source des données sur l'exposition aux antibiotiques dans les études cas-témoins évaluant les facteurs de risque d'antibiorésistance dans les bactériémies et les infections urinaires communautaires

	Bactériémies N = 15	Infections urinaires N = 32
Classification des antibiotiques		
<i>Exposition globale</i>	13	23
<i>Par famille / molécule</i>		
β-lactamines	3	2
Pénicillines	4	12
Pénicillines avec inhibiteurs des β-lactamases	3	6
Ampicilline ou ampicilline-sulbactam ou céfalotine	0	1
Mécillinam	0	1
Céfuroxime / C2G	3	5
Céphalosporines / C3G	4	17
Carbapénèmes	1	3
Aminoglycosides	1	6
Cotrimoxazole	0	10
Clindamycine	0	1
Doxycycline	0	1
(Fluoro)quinolones	7	20
Fosfomycine	0	3
Macrolides	1	5
Métronidazole	0	2
Nitrofurantoïne	1	7
Vancomycine	1	2
<i>Par spectre d'activité</i>		
Anti-anaérobies	1	0
Antibiotiques de spectre étroit / large	0	1
Pénicillines de spectre étroit / large	0	1
Source des données sur l'exposition		
Dossiers médicaux	8	19
Entretiens avec les patients	0	6
Délivrances dans les pharmacies	2	2
Évaluation prospective, sans détails	3	1
Non spécifiée*	2	4
Fenêtre d'exposition		
15 jours	1	0
1 mois	4	5
2 mois	1	0
3 mois	7	17
6 mois	1	4
12 mois	1	3
« Exposition récente »	0	1
31-365 jours avant	0	1
Non spécifiée*	0	1

* Certaines études ne précisait pas la fenêtre d'exposition [198] ou la méthodologie de recueil de l'exposition [162,179,180,183,198,206]. Abréviations : C2G : céphalosporines de 2^{ème} génération ; C3G : céphalosporines de 3^{ème} génération.

Définitions et impacts sur les résultats

Infections communautaires

Près de la moitié des études considéraient les infections communautaires vraies (n = 25/47), et l'autre moitié les infections communautaires (n = 20/47) ; de rares études se focalisaient sur les infections non hospitalisées (n = 1), ou communautaires/communautaires vraies (n = 1) sans les définir.

Les infections communautaires étaient généralement définies par leur survenue dans les 48 premières heures de l'hospitalisation (bactériémies, 9/11 ; IU, 3/9), excluant selon les études les patients transférés d'autres hôpitaux ou hospitalisés les dernières semaines (**Tableau 8.2**) [207]. D'autres définitions considéraient les patients ambulatoires, ceux se présentant au service des urgences, ou ceux adressés par leur médecin généraliste.

Les définitions des infections communautaires vraies (n = 25 études) n'étaient pas homogènes entre les études. Notamment, 13/25 études n'excluaient pas les infections associées aux soins ; le focus de ces études correspond plus aux infections communautaires qu'aux infections communautaires vraies. Une étude s'intéressant aux IU communautaires vraies ne les a pas définies.

L'exposition antérieure aux soins était analysée comme un facteur de risque d'antibiorésistance dans toutes les études se focalisant sur les infections communautaires. Étonnamment, des facteurs d'exposition aux soins étaient considérés dans 18/25 études se focalisant sur les infections communautaires vraies. Diverses définitions de l'exposition aux soins étaient utilisées, pouvant combiner plusieurs de ces critères : les critères de Friedman [67], les antécédents d'hospitalisation, de chirurgie, de procédure invasive des voies urinaires (avec différentes périodes avant l'admission), la résidence dans une maison de retraite médicalisée ou un établissement de soins de longue durée.

Tableau 8.2 Définitions des infections communautaires dans les études cas-témoins évaluant les facteurs de risque d'antibiorésistance dans les bactériémies et les infections urinaires

Bactériémies communautaires (N = 11)

< 48h (N = 4)

≤ 48h + pas de transfert (N = 3)

< 48h + pas de transfert + pas d'hospitalisation dans les 2 ou 4 semaines précédentes (N = 2)

Communautaires vraies ou résidence dans un ESLD (N = 1)

Admission aux urgences (N = 1)

Bactériémies communautaires vraies (N = 4)

≤ 48h (N = 1)

≤ 48h + non associées aux soins (N = 2)

Présentes ou en incubation à l'admission + pas d'hospitalisation dans le mois précédent (N = 1)

Infections urinaires communautaires (N = 9)

≤ 48h (N = 3)

<48h ou patient ambulatoire (N = 1)

Patient ambulatoire, ou diagnostic dans les 72h de l'admission aux urgences (N = 1)

Admission aux urgences (N = 1)

Admission aux urgences et pas d'hospitalisation ou de séjour dans un ESLD un mois avant (N = 2)

Identifiées dans un laboratoire de biologie + adressées par un généraliste (N = 1)

Infections urinaires non hospitalisées (N = 1)

Infections urinaires communautaires/communautaires vraies, sans définition (N = 1)

Infections urinaires communautaires vraies (N = 21)

Sans exclusion des infections associées aux soins (N = 11) : < 48h (N = 1), admission aux urgences (N = 3), prise en charge par un généraliste (N = 1), identifiées dans un laboratoire de biologie de ville (N = 1), patient ambulatoire (N = 5)

Avec exclusion des infections associées aux soins (N = 9) : pas d'autre critère (N = 3), admission aux urgences ou patients ambulatoires (N = 3), < 48h (N = 1), < 48h ou admission aux urgences (N = 1), ≤ 72h + pas d'hospitalisation dans les deux semaines précédentes (N = 1)

Non spécifiée (N = 1) [188]

ESLD : établissement de soins de longue durée

Cas-Témoins

Toutes les études de notre sélection définissaient le groupe cas comme des patients infectés par une bactérie résistante et incluaient un groupe témoin de patients infectés par une bactérie sensible (à l'antibiotique étudié) [151,152]. Les deux tiers des études se focalisaient sur les infections à E-BLSE (29/47 ; bactériémies, 8/29 ; IU, 21/29) ; 23 d'entre elles comprenaient une analyse multivariée (bactériémies, 7/23, **Tableau 8.3** ; IU, 16/23, **Tableau 8.4**). Ces dernières avaient comme groupe témoin des patients avec des infections à entérobactéries non BLSE [159,161–165,169,176–191]. L'exposition globale aux antibiotiques (OR univarié médian-min-max : 6,3-3,2-10,7 pour les bactériémies ; 4,3-2,0-10,1 pour les IU), en particulier aux céphalosporines (OR médian-min-max : 6,1-4,4-13,4 pour les bactériémies ; 3,5-1,4-17,1 pour les IU) et aux quinolones (OR médian-min-max : 6,4-3,0-7,0 pour les bactériémies ; 2,4-0,9-5,8 pour les IU), étaient des facteurs de risque en multivarié dans seulement la moitié des études (bactériémies, 3/4, 2/4 et 2/4 études ; IU, 6/11, 5/10 et 4/10 études, respectivement). Dans le cas des IU, l'exposition à d'autres antibiotiques, comme la pénicilline, les macrolides ou la nitrofurantoïne, constituait un facteur de risque dans une minorité d'études (2/8, 1/2, 1/4, respectivement). Les fenêtres d'exposition étudiées étaient variables mais la plupart étaient ≤3 mois. Une minorité d'études ont porté sur les infections communautaires vraies, avec un faible nombre de patients et une grande variabilité méthodologique, ne permettant pas de distinguer de spécificités propres à ce sous-groupe.

Aucune étude sur les bactériémies à E-BLSE n'a utilisé de témoins non infectés. Une seule étude sur les IU communautaires à E-BLSE a considéré deux groupes témoin parmi lesquels un groupe de patients non infectés, sélectionnés au hasard parmi les résidents de la même zone géographique que les cas [176].

Schéma cas-témoins

La plupart des études évaluées utilisaient un schéma cas-témoin, à l'exception de quatre études qui utilisaient un schéma cas-témoin(cas)-témoin. Ce schéma comprend un troisième groupe de patients, à considérer comme un deuxième groupe de cas (s'il est comparé au groupe témoin), ou un deuxième groupe de témoins (s'il est comparé au groupe de cas) [208]. Dans trois études, ce troisième groupe comprenait des patients présentant une septicémie [159], une bactériémie de toutes causes [161], ou

une infection par une bactérie différente des deux autres groupes [178]. Une seule étude a utilisé un schéma cas-témoins avec un deuxième groupe témoin de patients non infectés [176]. Ses résultats suggèrent que chez un individu non infecté issu de la population générale, l'exposition préalable à des antibiotiques à large spectre, au mécillinaam, au sulfaméthizole et au triméthoprime, pourrait augmenter le risque d'IU à *E. coli* non BLSE ou BLSE. D'autre part, l'exposition aux macrolides et à la nitrofurantoïne pourrait augmenter de manière sélective le risque d'infection urinaire à *E. coli* BLSE [176].

Tableau 8.3 Principaux résultats des études cas-témoins évaluant les facteurs de risque d'antibiorésistance dans les bactériémies communautaires à entérobactéries BLSE, selon le groupe témoin

Premier auteur, année	Groupe étudié	Nombre de cas	Nombre de témoins	Fenêtre d'exposition	AB identifié(s) comme Fdr	Odds ratio [IC95]	Autres AB évalués
Témoins = Entérobactéries non BLSE (N = 10)							
Kim, 2019 [163]	C	140	454	1 mois	β-lactamines ou FQ	6,8 [2,3-20,2]	
Lee, 2017 [164]	C	65	1 076	1 mois	Expo globale	15,3 [7,7-61,4]	
Park, 2011* [161]	C	50	100	3 mois	Expo globale	4 [1,5-12,3]	
Park, 2014† [162]	C	60	180	3 mois	C3G	16,4 [2,0-131,8]	C1G, C2G, Pénicillines, Quinolones
Rodríguez-Baño, 2010* [209]	C	95	188	2 mois	Expo globale FQ Céphalosporines	2,7 [1,5-4,9] 4,7 [2,0-11,1] 10,3 [2,1-50,3]	Aminopénicillines
Zahar, 2017 [210]	C vraies	22	335	1 an			β-lactamines, Céphalosporines, FQ, Carbapénèmes, Autres, Macrolides, Nitrofurantoïne, Aminoglycosides, Anti-anaérobies
Zahar, 2017 [210]	C AAS	36	289	1 an	Anti-anaérobies	3,51 [1,6-7,5]	β-lactamines, Céphalosporines, FQ, Carbapénèmes, Autres, Macrolides, Nitrofurantoïne, Aminoglycosides,
Gottesman, 2018 [169]	C vraies	12	267	3 mois	Quinolones	7 [1,7-29,4]	Exposition globale, Céfuroxime, Amoxicilline ou pénicilline, Amoxicilline-clavulanate
Témoins = sepsis communautaires (N = 1)							
Rodríguez-Baño, 2010 [209]	C	95	190	2 mois	Expo globale FQ	1,9 [1,03-3,5] 2,8 [1,2-6,5]	Aminopénicillines, Céphalosporines
Témoins = bactériémies communautaires (N = 1)							
Park, 2011* [161]	C	50	100	3 mois	Expo globale	3,9 [1,3-11,5]	

Une études sans analyse multivariée a été exclue [211]. * = schéma cas-cas(témoins)- témoins. † = pas d'information sur la méthodologie de recueil de l'exposition. Abréviations : AAS : associées aux soins ; AB : antibiotiques ; C : communautaires ; C1G, C2G, C3G : céphalosporines de 1^{ère}, 2^{ème}, 3^{ème} génération ; Expo globale : Exposition globale ; Fdr : facteur de risque ; FQ : fluoroquinolones.

Tableau 8.4 Principaux résultats des études cas-témoins évaluant les facteurs de risque d'antibiorésistance dans les infections urinaires communautaires à entérobactéries BLSE, selon le groupe témoin

Premier auteur, année	Groupe étudié	Nombre de cas	Nombre de témoins	Fenêtre d'exposition	AB identifié(s) comme Fdr	Odds ratio [IC95]	Autres AB évalués
Témoins = Entérobactéries non BLSE (N = 16)							
Søgaard, 2017* [212]	C	339	3 390	J-30 à -365	Macrolides Nitrofurantoïne	1,5 [1,1-2,2] 1,5 [1,1-2,3]	Mécillinam, Sulfaméthizole, TMP, Pénicillines à spectre large / étroit, AB à spectre large / étroit
Tüzün, 2019 [177]	C	154	151	6 mois	Expo globale	2,3 [1,2-4,3]	FQ, Céphalosporines, β -lactamines + inhibiteurs des β -lactamases
Zhu, 2019 [181]	C vraies (en réalité C)	111	103	3 mois	Expo globale	4,2 [1,9-9,1]	
Boix-Palop, 2017* [178]	C	83	155	3 mois	Céphalosporines Quinolones	4,0 [1,8-9,2] 3,3 [1,4-7,9]	
Kung, 2015+ [179]	C	62	376	3 mois			Pénicillines avec / sans inhibiteurs des β -lactamases, Céphalosporines, FQ
Topaloglu, 2010 [182]	C vraies (en réalité C)	155	155	3 mois			Exposition globale, Pénicillines, C2G, C3G+C4G, Autres, Combinaisons, AB suppressive par cotrimoxazole ou Nitrofurantoïne
Yilmaz, 2008+ [183]	C / C vraies	62	62	3 mois	Quinolones Céphalosporines	NA NA	Exposition globale, Aminopénicillines, Aminoglycosides, Aminopénicillines + inhibiteurs des β -lactamases, Cotrimoxazole
Colodner, 2004 [184]	Non hospitalisées	100	77	3 mois	Expo globale C2G C3G Quinolones Pénicillines	3,2 [1,8-5,9] 15,8 [1,7-143] 10,1 [4,2-24,0] 4,1 [1,8-9,0] 4,0 [1,6-9,0]	Macrolides, C1G, Cotrimoxazole, Nitrofurantoïne, Aminoglycosides
Calbo, 2006+ [180]	C (en réalité C vraies)	19	55	1 an	Céfuroxime	21,4 [5,4-85,2]	C3G, Ciprofloxacine, Aminoglycosides
Albaramki, 2019 [185]	C vraies	77	77	3 mois			Expo globale

Guzmán, 2019 [186]	C vraies	21	82	3 mois			« β-lactamines de 1 ^{ère} ligne » (Ampicilline, Ampicilline-sulbactam, Céfaloine)
Koksal, 2019 [187]	C vraies	66	112	3 mois			Expo globale, Pénicillines, Céphalosporines, FQ, Fosfomycine, Nitrofurantoïne
Azap, 2010 [‡] [188]	C vraies	51	464	3 mois	β-lactamines	4,6 [2,0–10,7]	Expo globale, Quinolones
Rodríguez-Baño, 2008 [189]	C vraies	122 (93% avec IU)	244	3 mois	Expo globale Aminopénicillines FQ C2G + C3G	2,9 [1,6-5,3] 2,8 [1,2-6,1] 1,9 [1,0-4,1] 3,2 [1,0-9,5]	Fosfomycine, Trométhamine
Castillo-Tokumori, 2017 [190]	C vraies	67	105	1 mois	Expo globale	3,1 [1,4–6,7]	Pénicillines, Céphalosporines, FQ
Megged, 2014 [191]	C vraies	80	80	1 mois	Expo globale	4 [1,6–10,4]	
Témoins = patients non infectés de la population générale (N = 1)							
Søgaard, 2017* [212]	C	339	3 390	J -30 à -365	Macrolides Nitrofurantoïne AB large spectre Mécillinam Sulfaméthizole TMP	2,2 [1,3-3,8] 4,7 [1,9-11,2] 2,9 [1,1-7,7] 3,1 [1,3-7,8] 4,7 [2,4-9,2] 3,6 [1,6-8,0]	Pénicillines à spectre étroit / large, AB à spectre étroit



Abréviations : AB : antibiotiques ; C : communautaires ; C1G, C2G, C3G, C4G : céphalosporines de 1^{ère}, 2^{ème}, 3^{ème}, 4^{ème} génération ; Expo globale : Exposition globale ; Fdr : facteur de risque ; FQ : fluoroquinolones ; TMP = triméthoprim. Les études sans analyses multivariées ont été exclues [174,175,192–194]. * = schéma cas-cas(témoins)- témoins. † = pas d'information sur la méthodologie de recueil de l'exposition, ‡ = pas de définition des infections communautaires vraies.

8.4 Discussion

La revue de 47 études étudiant l'exposition aux antibiotiques comme facteur de risque d'antibiorésistance dans les infections urinaires ou les bactériémies communautaires a permis de souligner leur hétérogénéité méthodologique et son impact sur les résultats.

Discussion des principaux résultats

Facteur de risque : l'exposition aux antibiotiques

Les antibiotiques étaient classifiés de manière hétérogène selon les études. Trois études sur quatre considéraient l'exposition globale aux antibiotiques, et la plupart considéraient également les antibiotiques en familles ou par molécule. L'étude globale, ou par molécule / famille, sont des approches complémentaires. En effet, la première considère l'impact global de l'exposition aux antibiotiques sur le processus menant à l'antibiorésistance, la deuxième évalue le lien entre un(e) (famille d') antibiotique(s) et le développement de l'antibiorésistance. Dans les cas où des comparaisons multiples sont effectuées, les p-value devraient être corrigées. Seule une étude sur quatre a individualisé les antibiotiques à spectre anti-anaérobie. Compte tenu du rôle des anaérobies en tant que flore barrière [150,213], il semble instructif d'isoler les anti-anaérobies en tant que groupe distinct dans les futures études.

Les données sur l'exposition aux antibiotiques étaient majoritairement recueillies dans les dossiers médicaux, ce qui peut induire un biais de représentativité car l'exposition aux antibiotiques n'y est pas recueillie et retracée de manière systématique et homogène, et sous-estimer l'estimation de l'exposition. Les autres sources pouvaient être les entretiens avec les patients, qui sont sujets à un biais de rappel, ou les bases de données de pharmacie. Les données des pharmacies en communauté concernent souvent la délivrance d'antibiotiques, qui peut ne pas refléter la consommation réelle des patients et donc ici aussi induire un biais si la définition de l'exposition s'appuie sur des mesures quantitatives. Néanmoins, du fait de leur exhaustivité, les données des pharmacies semblent être une source de donnée optimale. Les fenêtres d'exposition aux antibiotiques étaient majoritairement ≤ 3 mois ; seule une étude sur 5 considérait une fenêtre d'exposition dépassant 3 mois. Le choix de la fenêtre d'exposition est essentiel, en particulier lorsqu'il s'agit de bactéries qui peuvent s'associer à un portage chronique. Par exemple, la durée du portage intestinal des E-BLSE peut varier considérablement (et parfois dépasser 12 mois) en fonction du contexte d'acquisition, de la population, et de la souche [146,147]. Ainsi, la définition de la fenêtre d'exposition doit s'appuyer sur les informations disponibles sur la dynamique d'acquisition et de perte de la paire bactérie-résistance considérée, et sur la population étudiée. Par ailleurs, si l'on considère les antibiotiques trop près de l'hospitalisation, ils seront potentiellement associés à l'infection, causant un biais protopathique. Ainsi, la fenêtre d'exposition devrait exclure l'exposition récente aux

antibiotiques ; seule une étude le mentionnait explicitement [176]. Le choix de la période d'exposition récente à exclure peut être difficile et dépend de l'infection et du contexte de l'étude.

De manière plus globale, dans de futures études, une analyse de sensibilité testant plusieurs fenêtres d'exposition devrait être envisagée. A titre d'exemple, Opatowski et al. ont testé plusieurs fenêtres d'exposition pour étudier les facteurs de risque d'infection urinaire communautaire résistante aux antibiotiques à partir des données du SNDS [200]. L'étude a montré que les expositions au-delà de 90 jours n'avaient pas d'effet en univarié sur la résistance, ce qui justifiait le choix de cette fenêtre d'exposition dans le modèle.

Infections communautaires

Notre analyse a montré une grande labilité entre les définitions des infections communautaires et des infections communautaires vraies, avec une variété importante de définitions utilisées pour les infections associées aux soins. Au vu de ces résultats, on peut s'interroger sur l'intérêt de différencier les infections communautaires vraies des infections communautaires. En effet, cette catégorisation est difficile, car : (1) l'identification des infections communautaires vraies nécessite d'isoler les infections associées aux soins, dont la définition n'est pas standardisée et dépend des infections [88,214,215]; (2) l'obtention d'une information exhaustive sur les parcours de soins des patients avant l'hospitalisation est difficile.

Par ailleurs, l'intérêt de cette catégorisation pour la prise en charge des patients, notamment dans le choix du traitement empirique, est discutée [216]. Elle reposait initialement sur des travaux suggérant un risque plus élevé d'antibiorésistance et de mortalité dans les infections communautaires associées aux soins, justifiant des mesures de diagnostic et de traitement agressives [207,214,217]. Cependant, cette catégorisation a été abandonnée dans les recommandations américaines de prise en charge des pneumonies communautaires [218,219]. Dans ces recommandations, l'accent est mis sur l'épidémiologie locale et les facteurs de risque validés pour déterminer la nécessité de couvrir le SARM ou le *P. aeruginosa* lors du choix du traitement empirique dans les pneumonies communautaires. En effet, le risque d'antibiorésistance et de mortalité attribuable aux pneumonies communautaires associées aux soins semble avoir été surestimé et entraîner des excès thérapeutiques [216]. En outre, les travaux évaluant l'impact d'un traitement empirique inadapté sur la mortalité de patients ayant une bactériémie communautaire montrent des résultats discordants [159,161].

Du fait de la persistance d'inconnues, il semble intéressant d'un point de vue épidémiologique de continuer à isoler les facteurs d'exposition aux soins dans de futures études, afin de mieux définir les populations à risque. Par exemple, une étude de Lin et al. a identifié l'exposition aux antibiotiques comme un facteur de risque de bactériémie communautaire à E-BLSE, mais pas dans le sous-groupe de bactériémies communautaires vraies [23]. Ces résultats pourraient suggérer que l'effet de l'exposition aux antibiotiques sur la résistance est restreint aux patients avec exposition aux soins. Enfin, s'il semble

difficile de parvenir à un consensus sur une définition opérationnelle des infections communautaires associées aux soins, les études cas-témoins sur les infections communautaires devraient préciser clairement les définitions utilisées, et exclure les patients ayant été exposés à des soins du groupe communautaire vrai.

Témoins

La définition des témoins est une étape cruciale, qui dépend de la question scientifique. La plupart des études cas-témoins dans la littérature, et dans notre sélection, définissent le groupe témoin comme des patients infectés par une bactérie sensible à l'antibiotique étudié [151,152]. Cela permet de comparer l'impact différentiel de l'exposition aux antibiotiques sur la survenue d'une infection sensible ou résistante aux antibiotiques. Pour les cliniciens, ces études aident à définir le traitement empirique optimal devant une suspicion d'infection bactérienne (par exemple, déterminer si des carbapénèmes sont justifiées pour couvrir les souches d'E-BLSE lors du choix du traitement empirique devant une IU) [151,208]. Alors qu'elles sont généralement considérées comme des facteurs de risque d'E-BLSE, les études analysées dans notre sélection montraient que l'exposition globale aux antibiotiques, en particulier aux céphalosporines et aux quinolones, étaient des facteurs de risque dans seulement la moitié des cas. Des différences dans les populations, la taille, le contexte des études, l'hétérogénéité dans la définition de l'exposition et des cas, pourraient avoir contribué à la variabilité de ces résultats. Particulièrement, les résultats de cette revue ne permettent pas de trancher sur l'impact de l'exposition aux antibiotiques dans les infections communautaires vraies sur le risque de résistance. Les fenêtres d'exposition étaient majoritairement ≤ 3 mois, empêchant toute conclusion sur l'impact d'expositions plus anciennes.

Dans notre sélection, une seule étude considérait des témoins non infectés. Pour étudier les infections communautaires, il semble important d'inclure dans de futures études un groupe de témoins non infectés. En effet, cela permettrait d'identifier l'impact de l'exposition aux antibiotiques sur le processus global conduisant les individus de la communauté à s'infecter. Cependant, l'inclusion de témoins non infectés peut s'avérer très difficile, notamment lorsqu'il ne s'agit pas d'une population spécifique [151–154]. De plus, les témoins non infectés doivent remplir les mêmes critères de non-exposition aux soins que ceux choisis par les auteurs pour définir l'infection communautaire (par exemple, ne pas avoir été hospitalisés au cours de la semaine précédant l'inclusion, etc.). Enfin, selon le pathogène étudié et la question scientifique, ce groupe pourrait être la population générale (si la bactérie est commensale), ou des individus non colonisés issus de la population générale pour les bactéries non commensales.

Schéma cas-témoins

Parmi les 47 études analysées, une seule utilisait un schéma cas-témoin(cas)-témoin avec un premier groupe de témoins infectés avec la bactérie sensible (à l'antibiotique d'intérêt), et un deuxième groupe de témoins non infectés. Pourtant, un tel schéma permettrait de déterminer si l'administration d'antibiotiques à des individus de la communauté a un impact sélectif sur le risque de développer une infection résistante ou sur le risque d'avoir une infection, qu'elle soit sensible ou résistante [208]. En effet, la comparaison des patients non infectés avec les deux autres groupes identifierait de manière séparée l'impact de l'exposition aux antibiotiques sur la survenue d'une infection sensible, ou résistante aux antibiotiques.

Proposition : hypothèses physiopathologiques

Afin d'améliorer la standardisation des études futures, nous proposons de les construire et de les interpréter dans le cadre d'une hypothèse physiopathologique (c'est-à-dire causale) spécifique à la paire bactérie-résistance d'intérêt (**Tableau 8.5**). Cette hypothèse détaillerait les étapes conduisant un hôte non colonisé / non infecté de la communauté à présenter une infection résistante aux antibiotiques. Les auteurs pourraient intégrer leur étude dans ces étapes. L'hypothèse pourrait être formulée à un niveau macro (qui fait intervenir des groupes de patients, pour aider à concevoir des études cas-témoins), ou à un niveau micro (qui considère les interactions et les dynamiques intra-hôte et hôte-microbiote, pour d'autres types d'études). Cette stratégie aiderait à concevoir l'étude de manière à apporter la réponse la plus correcte à la question spécifique posée. En groupant le résultat des études, cela permettrait d'avoir une vision globale du rôle des antibiotiques à chaque étape menant à une infection résistante aux antibiotiques, pour une paire bactérie-résistance spécifique. Cela permettrait également d'identifier les transitions pour lesquelles l'impact de l'exposition aux antibiotiques n'a pas été étudié, incitant à réaliser des études supplémentaires.



Tableau 8.5 Points clés pour concevoir des études cas-témoins évaluant l'exposition aux antibiotiques comme facteur de risque d'infections communautaires résistantes aux antibiotiques.

Identifier le couple bactérie-résistance d'intérêt.

Définir l'hypothèse physiopathologique, définissant la population source non colonisée / non infectée hypothétique et les étapes menant à l'infection.

Déterminer la ou les étapes, et la question spécifique, que l'étude veut explorer.

Choisir les groupes cas et témoins en conséquence.

Construction des hypothèses physiopathologiques

L'hypothèse physiopathologique détaille la chaîne causale hypothétique conduisant un hôte non colonisé / non infecté à présenter une infection par la bactérie résistante. Pour qu'une infection survienne chez un individu, il faut généralement que celui-ci soit préalablement colonisé par cette bactérie [150]. Selon la paire bactérie-résistance étudiée, la bactérie résistante peut être acquise directement (comme dans le cas de la résistance à la pénicilline de *S. aureus*), ou il peut être nécessaire de passer par une étape de colonisation par une souche sensible acquérant secondairement une résistance (comme on pourrait le supposer pour *P. aeruginosa*). Par la suite, l'infection se différencie de la colonisation par la présence de signes cliniques signalant le passage du portage asymptomatique à l'infection [12].

En fonction de la paire bactérie-résistance, une hypothèse physiopathologique au niveau macro pourrait inclure quatre états, qui pourraient chacun s'incarner par un groupes de patients spécifiques : (1) absence de portage de la bactérie ; (2) portage de la bactérie sensible (par ex, colonisation par une nouvelle souche d'*Escherichia coli*, colonisation transitoire par *Streptococcus pneumoniae* ou *S. aureus*, colonisation transitoire ou persistante par *Enterobacter cloacae*, *Serratia spp.*, *Klebsiella spp.*) ; (3) portage de la bactérie résistante (par mutation d'une bactérie sensible, ou transfert latéral de gènes, ou acquisition de la résistance à partir d'un autre hôte) ; (4) infection par la bactérie résistante [12,157,220]. Lors de la conception d'une étude cas-témoins, la clarification de la question scientifique aiderait à choisir les groupes cas et témoins les plus appropriés (**Tableau 8.6**).

Tableau 8.6 Sélection des groupes cas et témoins en fonction de la question spécifique

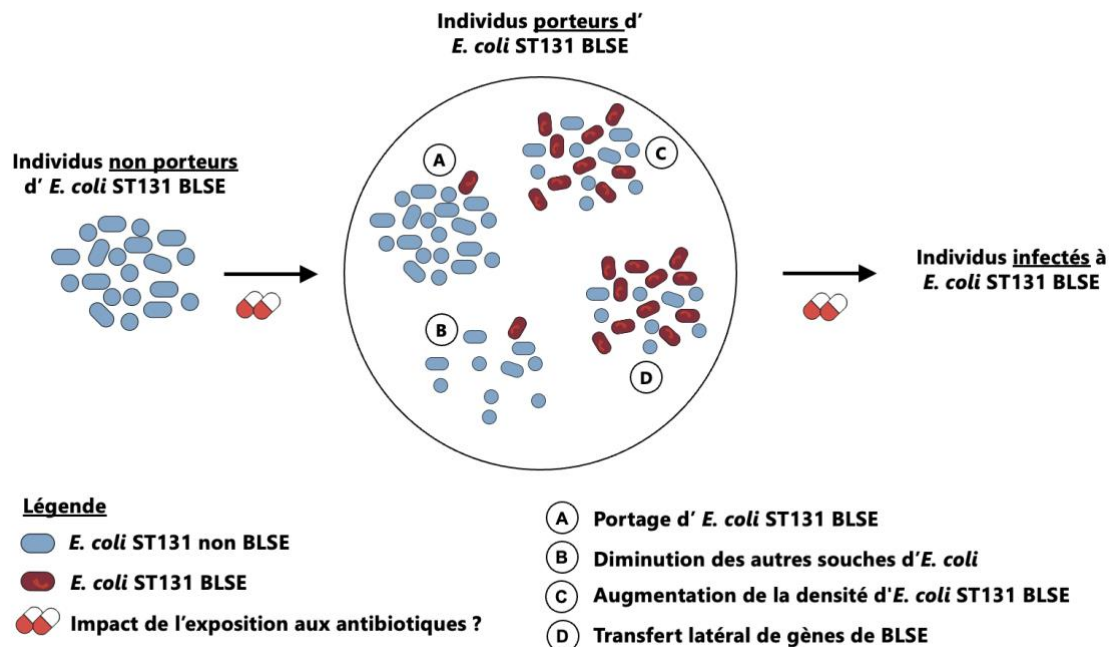
Témoins	Cas	Question pour laquelle l'impact de l'exposition est évalué
Hôtes non colonisés	Hôtes colonisés par une espèce / souche donnée	Colonisation
Hôtes colonisés par la bactérie sensible	Hôtes colonisés par la bactérie résistante	Apparition d'une souche résistante parmi une souche colonisatrice
Hôtes colonisés par une souche bactérienne	Hôtes infectés par la même souche bactérienne	Progression de la colonisation vers l'infection
Hôtes non infectés / non colonisés	Hôtes infectés par la bactérie résistante	Progression de la non-infection / non-colonisation à l'infection résistante
Témoins 1 : hôtes non infectés ; témoins 2 : hôtes infectés par la bactérie sensible	Hôtes infectés par la bactérie résistante	Impact différentiel de l'exposition sur le risque de développer une infection résistante, ou d'avoir une infection, qu'elle soit sensible ou résistante
Hôtes infectés par la bactérie sensible	Hôtes infectés par la bactérie résistante	Impact différentiel de l'exposition sur le risque de développer une infection sensible ou résistante.

*Le groupe témoins 2 pourrait être considéré comme un groupe cas 2, selon les comparaisons statistiques réalisées dans l'étude.

Un exemple : infection communautaire à *E. coli* ST131 BLSE

Prenons l'exemple d'*E. coli* ST131 BLSE, une souche multirésistante pandémique qui se répand dans la communauté, pour illustrer l'interaction entre l'hypothèse physiopathologique et la conception des études cas-témoins [221]. Pour étudier les facteurs de risque d'infection à *E. coli* ST131 BLSE, une hypothèse physiopathologique en 4 étapes, partant de la population générale comme population source [221], peut être proposée (**Figure 8.2**) : (1) absence de portage d'*E. coli* ST131 ; (2) portage transitoire d'*E. coli* ST131, après acquisition par transmission inter-individuelle ; (3) portage d'*E. coli* ST131 BLSE, après acquisition d'un plasmide avec un gène codant une BLSE, par transmission horizontale entre bactéries ; (4) infection à *E. coli* ST131 BLSE, par exemple après translocation digestive [221].

Figure 8.2 Hypothèse physiopathologique pour étudier l'exposition aux antibiotiques en tant que facteur de risque d'infection par *Escherichia coli* ST131 BLSE



Pour comprendre et évaluer le rôle potentiel des antibiotiques dans le passage d'une étape à l'autre, un schéma spécifique est nécessaire pour chaque transition.

Étape 1 à 2 : Morales Barroso et al. ont étudié le rôle de l'exposition aux antibiotiques dans l'acquisition d'*E. coli* ST131 (indépendamment de la production de BLSE) chez un hôte non infecté. Ils ont comparé des individus porteurs et non porteurs d'*E. coli* ST131, appartenant tous aux foyers de patients infectés à *E. coli* ST131. L'exposition aux antibiotiques n'était pas un facteur de risque de portage [222].

Étape 2 à 3 : à notre connaissance, aucune étude cas-témoins n'a comparé des individus porteurs d'*E. coli* ST131 non BLSE à des individus porteurs d'*E. coli* ST131 BLSE.

Étape 3 à 4 : pour évaluer l'impact de l'exposition aux antibiotiques dans la transition entre la colonisation et l'infection par *E. coli* ST131 BLSE, une étude optimale devrait inclure des cas infectés par *E. coli* ST131 BLSE et des témoins non infectés porteurs d'*E. coli* ST131 BLSE. A notre connaissance, une telle étude n'a pas été publiée. Plus généralement, s'il a été démontré que l'exposition aux antibiotiques a un impact sur le risque de colonisation par certaines entérobactéries BMR, leur rôle dans la progression de la colonisation vers l'infection est partiellement connu [150]. Les études sur ce sujet s'intéressaient en général aux patients hospitalisés, en particulier dans les unités de soins intensifs, où la colonisation par des bactéries BMR est souvent recherchée en pratique courante [223]. Pour mieux répondre à cette question, il est important de concevoir des études spécifiques en milieu communautaire avec des individus colonisés comme témoins et des individus infectés comme cas [224–227].

Certaines études ont comparé des cas avec des infections à *E. coli* ST131 BLSE et des témoins avec des infections à *E. coli* non ST131 BLSE [228,229]. Ce schéma examine spécifiquement l'impact de l'exposition aux antibiotiques sur le risque de développer une infection due à la souche ST131, parmi les patients qui développeraient des infections à *E. coli* BLSE.

Si les études cas-témoins évaluent les facteurs de risque impliqués dans la transition d'une étape à l'autre de l'hypothèse physiopathologique, elles ne distinguent pas les interactions et les dynamiques intra-hôte et hôte-microbiote impliquant l'antibiotique dans cette transition, qui sont mieux abordées par des modèles mathématiques et nécessitent une analyse du microbiome [151,230,231]. Dans l'exemple présenté ci-dessus, une hypothèse physiopathologique au niveau micro pourrait impliquer d'autres étapes entre le portage et l'infection à *E. coli* ST131 BLSE (**Figure 8.2**) : (1) diminution des autres souches d'*E. coli* dans le microbiote de l'hôte [150] ; (2) augmentation du transfert latéral de gènes de BLSE au sein d'une espèce bactérienne ou entre espèces bactériennes [230] ; (3) augmentation de la densité d'*E. coli* ST131 BLSE dans le microbiote de l'hôte [150] (une étape qui pourrait être importante pour les bactériémies [232]), avec une perte de diversité bactérienne. Toutes ces étapes pourraient se produire au sein du microbiote des sujets porteurs d'*E. coli* ST131 BLSE et être influencées par la pression sélective des antibiotiques.

8.5 Apport de la base BactHub

Dans ce travail, nous avons souligné l'importance de clarifier le lien entre l'exposition individuelle aux antibiotiques et l'occurrence d'infections bactériennes communautaires résistantes aux antibiotiques. La clarification de ce lien doit avoir lieu sur toutes les étapes menant un individu sain à s'infecter, et dans les infections communautaires vraies, qui concernent la population générale.

En l'état, pour une paire bactérie-résistance donnée, les données de la base BactHub permettent d'effectuer des études cas témoins comparant des patients infectés avec la bactérie résistante ou sensible. Des études avec des témoins porteurs d'une bactérie sensible sont également possibles, mais s'accompagneraient d'un biais de sélection car la colonisation n'est recherchée que chez certains patients ou dans des contextes particuliers en pratique clinique. L'étude de l'exposition aux antibiotiques avant l'hospitalisation comme facteur de risque est impossible, car les données se limitent aux séjours hospitaliers des patients. Enfin, l'absence de données sur l'exposition antérieure des patients aux soins ne permet pas d'isoler les épisodes communautaires vrais parmi l'ensemble des épisodes communautaires.

Pour l'étude du lien entre l'exposition aux antibiotiques et la survenue de bactériémies communautaires résistantes, plusieurs enrichissements de la base BactHub sont nécessaires : (1) un groupe de témoins non infectés doit être constitué, avec des informations cliniques détaillées, et (2) des données sur l'exposition antérieure aux antibiotiques et aux soins des patients hospitalisés avec une bactériémie doivent être recueillies.

Pour construire un groupe de témoins non infectés, nous avons identifié au sein de l'EDS de l'AP-HP tous les patients majeurs avec (1) au moins une hospitalisation complète de durée < 6 mois entre le 1^{er} janvier 2017 et le 31 décembre 2019, et (2) aucun prélèvement microbiologique positif (à toute bactérie ou champignon), et (3) aucun code cim-10 d'infection aigue ou chronique à bactérie ou à mycobactérie (**Annexe 14**). Puis, nous avons apparié chaque séjour avec un épisode de bactériémie communautaire entre le 1^{er} janvier 2017 et le 31 décembre 2019 de la base BactHub, avec cinq séjours de patients uniques non infectés sur l'âge, le sexe \pm 5 ans, l'année, le mois et la période d'hospitalisation (avril-septembre vs octobre-mars)²⁹. Les données des témoins ont été structurées de la même manière que celles des patients hospitalisés avec une bactériémie, mais à l'échelle du séjour, sans faire intervenir la notion d'épisode.

Des témoins non hospitalisés auraient également pu être identifiés au sein du SNDS. Cependant, les délais du projet ne permettaient pas de faire une telle demande auprès de la CNIL au regard du temps

²⁹ L'appariement n'a pas pu se faire sur l'hôpital, car le numéro FINESS permettant de l'identifier n'a été mis à disposition qu'en février 2021, après réalisation de l'appariement.

mis par la Cnam pour la livraison de telles données. C'est pourquoi nous avons préféré demander seulement un appariement indirect avec le SNDS, en limitant les données extraites.

Cet enrichissement de la base BactHub avec des données du SNDS permettrait d'avoir des données sur l'exposition antérieure aux antibiotiques et aux soins des patients. La pandémie liée à la covid-19, la multiplicité des acteurs impliqués dans le projet (INSERM, AP-HP, HDH), les délais pour l'obtention des documents nécessaires pour déposer une demande d'autorisation à la CNIL et l'utilisation de la plateforme du HDH, ont allongé les délais d'obtention de l'autorisation CNIL nécessaire pour réaliser l'appariement. Disposant des informations du PMSI dans les deux sources, le taux d'appariement attendu était proche de 100%. L'étude du lien entre l'exposition aux antibiotiques et la survenue d'une bactériémie communautaire résistante était l'objectif initial de cette thèse, et avait été formalisée dans un projet lauréat de l'appel à projets du HDH en janvier 2019. Elle n'a pas pu être réalisée dans le cadre de cette thèse, mais pourra faire l'objet de travaux futurs, l'autorisation CNIL ayant été obtenue le 22 juillet 2022.



9 DISCUSSION GÉNÉRALE

« Les fils s'assemblent, prennent forme, se mêlent, se nouent, parfois ils se démêlent puis se renouent. C'est ça, le fil du temps »

Hitoha Miyamizu, *Your Name*, Makoto Shinkai

9.1 Synthèse des principaux résultats

Cette thèse avait pour objectif d'évaluer l'apport des données hospitalières pour l'étude des infections bactériennes résistantes aux antibiotiques, à travers l'exemple des bactériémies et de l'EDS de l'AP-HP. Pour répondre à cet objectif, nous avons présenté les travaux de structuration de la base BactHub de patients hospitalisés avec une bactériémie. Cette base est la première en France à présenter des données détaillées cliniques et microbiologiques pour plus de 30 000 patients hospitalisés avec une bactériémie dans 14 hôpitaux universitaires de l'AP-HP avec une activité de soins aigus entre 2016 et 2019. Les hôpitaux inclus couvrent plus de 75% des lits de soins aigus de l'AP-HP, et environ 22% des séjours de soins aigus dans la région île de France. Les données structurées incluent des informations sur les patients (sexe, âge, date de décès (le cas échéant), comorbidités détaillées, etc.), et sur l'ensemble de leurs séjours de 2016 à 2019, y compris leur séjour d'inclusion (date d'admission et de sortie, passages dans les UH, codages CIM-10, actes de la CCAM, dates et résultats des prélèvements biologiques et microbiologiques, date et résultats d'une partie des observations physiologiques et une partie des données médicamenteuses, comptes-rendus d'hospitalisation, etc.). Malgré la diversité et la richesse des données incluses, certains écueils ont été soulignés, comme la mauvaise complétude des données d'exposition aux antibiotiques à l'hôpital, ou l'impossibilité de calculer le score de sévérité SOFA de l'infection.

La détermination du site primaire des bactériémies est un point important de leur caractérisation. Cette détermination est compliquée du fait du nombre important d'infections bactériennes à définir, et de la présence incomplète de certaines données physiologiques, biologiques, ou radiologiques nécessaires à leur diagnostic dans la base BactHub. Les codages CIM-10 sont utilisés dans la base pour définir les sites primaires, mais la fiabilité de cette technique n'est pas démontrée dans ce cadre d'application. Aussi, un travail exploratoire d'identification des sites primaires des bactériémies dans les CRH par TAL a été présenté. Son apport et celui des codages CIM-10 ont été confrontés à une lecture experte des CRH. La proportion de sites primaires réels parmi ceux identifiés était deux fois plus élevée avec le programme

de TAL qu'avec les codages CIM-10. Les CRH de plus d'un épisode sur 4 ne contenaient pas d'information sur l'épisode infectieux ; par conséquent, la part des sites primaires identifiés parmi la totalité des sites primaires était 1,4 fois plus élevée avec les codages CIM-10. Si aucune technique n'est parfaite à elle seule, l'interprétation des résultats du programme de TAL doit tenir compte de la fréquente présence de faux négatifs, et ceux des codages CIM-10 de la fréquente présence de faux positifs.

Par la suite, nous avons décrit les principales caractéristiques des patients, des épisodes de bactériémie, et des isolats bactériens dans la base, selon le caractère communautaire ou nosocomial de l'épisode. Les patients inclus avaient pour la plupart des comorbidités, et présentaient un fort taux de mortalité atteignant plus de 20% à J90. Les taux de mortalité associés aux épisodes communautaires sont plus faibles (~5%) qu'ils soient intra-hospitalier, à J30 ou J90. Les épisodes communautaires présentaient une durée médiane deux fois plus courte, étaient moins souvent polymicrobiens, avec moins de passage en soins intensifs, et moins de bactériémies primaires. Les isolats bactériens présentaient de forts taux d'antibiorésistance dans les deux groupes (*K. pneumoniae* C3G-R, 21-37% ; *E. coli* C3G-R 13-17% ; SARM 11-14%), qui variait selon les sites primaires. L'incidence qui a été estimée semble élevée (20-22 séjours avec bactériémie / 1 000 séjours) et augmente de 2017 à 2019. Ces résultats ont été confrontés aux données de la littérature et aux données de surveillance de l'antibiorésistance, confortant l'utilisation de la base pour la recherche. A titre d'exemple, les données de la base pourraient permettre d'étudier les parcours de soins des patients avec bactériémie pendant leur séjour à l'hôpital, ou de construire des études cas-témoins pour évaluer des facteurs de risque d'antibiorésistance, de mortalité, de réadmission dans un hôpital de l'AP-HP (jusqu'à 1 an après le séjour avec bactériémie), de rechute, ou de récurrence de bactériémie.

Nous avons ensuite étudié les facteurs de risque de récurrence de bactériémie à 1 an sur plus de 3 500 épisodes incidents de bactériémie communautaire. Dans ce travail, nous montrons pour la première fois que la résistance aux C3G de *Klebsiella spp* et d'*E. coli* dans les bactériémies communautaires est un facteur de risque de récurrence de bactériémie à 1 an. Nous décrivons un impact différentiel selon le germe ; l'odds ratio associé à la récurrence était deux fois plus élevé pour *Klebsiella spp* C3G-R (OR 4,7) que pour *E. coli* C3G-R (OR 2,5), alors que l'identification d'un SARM n'était pas liée à la récurrence. Notre étude a permis également de préciser les comorbidités et les sites primaires associés à la récurrence dans le cadre communautaire. Ainsi, nous avons montré que la présence d'un cancer, et secondairement d'une maladie hépatique ou rénale, était associée à la récurrence. De même, l'absence de site primaire, un site primaire digestif ou sur matériel augmentait le risque de récurrence. Nos résultats inédits incitent à une vigilance accrue vis-à-vis des infections à *Klebsiella spp*, au vu de la diffusion des BLSE en communauté, et la production de BLSE étant le principal mécanisme de résistance aux C3G.

Enfin, l'importance de clarifier le lien entre l'exposition individuelle aux antibiotiques et l'occurrence d'infections bactériennes communautaires résistantes aux antibiotiques a été soulignée. A travers une revue des travaux dédiés dans les infections urinaires et les bactériémies, j'ai présenté l'hétérogénéité existante en rapport avec l'évaluation de l'exposition aux antibiotiques, la définition des cas, des témoins, et le schéma de l'étude. L'impact de cette hétérogénéité sur les résultats a été discutée, avec l'exemple des E-BLSE. Notamment, j'ai montré que les expositions globales aux antibiotiques, aux céphalosporines ou aux quinolones n'étaient des facteurs de risque d'E-BLSE que dans la moitié des études les évaluant. Par ailleurs, la paucité d'études avec des témoins non infectés a été mise en avant, alors que ceux-ci sont essentiels pour approcher la population source de l'infection. A cela s'ajoute le faible nombre de travaux sur les infections communautaires vraies, avec une labilité importante entre les infections communautaires vraies et les infections communautaires associées aux soins. Par la suite, j'ai présenté les perspectives d'utilisation de la base BactHub pour étudier le lien entre l'exposition aux antibiotiques et l'occurrence d'infections bactériennes communautaires résistantes. L'autorisation de la CNIL a été obtenue le 22 juillet 2022, pour enrichir la base avec des données issues du SNDS, permettant d'isoler les bactériémies communautaires vraies et de reconstituer l'exposition aux antibiotiques des témoins et des patients avant leur bactériémie.

9.2 Les EDS en épidémiologie

Les travaux de cette thèse se sont appuyés sur l'EDS de l'AP-HP, permettant de mettre en évidence les avantages et les défis liés à l'utilisation des EDS pour la recherche en épidémiologie.

Avantages



Les EDS donnent accès à des données de vie réelle, recueillies dans le cadre de la pratique clinique courante. Ces données permettent d'approcher la vraie population concernée par la maladie ou la thérapeutique étudiée à l'échelle d'un ou de plusieurs hôpitaux, à la différence des données relevant d'études interventionnelles où la population d'étude est définie par des critères d'inclusion stricts, parfois éloignés de la vraie vie et pouvant exclure les populations aux âges extrêmes ou avec des comorbidités. Les données des EDS permettent également de réaliser des études en situation de vie réelle, prenant en compte les modalités d'accès et de pratique du soin dans la vraie vie.

Par ailleurs, les EDS sont créés à l'échelle d'un ou de plusieurs établissements de santé ; ils peuvent couvrir de manière quasi-exhaustive la population prise en charge dans ces établissements. Aussi, ils permettent de réaliser des études sur un grand volume de patients, et ce en s'abstenant des contraintes

humaines, financières et temporelles importantes que nécessitent les études impliquant la personne humaine, ou les études rétrospectives nécessitant un retour aux dossiers médicaux.

Les EDS contiennent une grande diversité d'informations hospitalières cliniques et paracliniques, dans le périmètre des hôpitaux de chaque entrepôt. Notamment, ils contiennent des données physiologiques, radiologiques, et d'examen physique absentes des données des réseaux de surveillance et des bases de données médico-administratives, ainsi que des résultats biologiques et des données médicamenteuses à l'hôpital, non recueillis dans les bases de données médico-administratives.

Enfin, la réutilisation des données des DPI pour la recherche est potentialisée par la mise en place d'EDS, dont le développement devrait croître, comme l'atteste le nombre important d'EDS mis en place en France ces dix dernières années. Le ministre des solidarités et de la santé et le secrétaire général pour l'investissement et pilote du plan France 2030 ont annoncé en mars 2022 le lancement prochain d'un appel à projets doté de 50 millions d'euros pour accompagner et soutenir la constitution d'EDS coordonnées avec le HDH, dans le cadre de la stratégie d'accélération en santé numérique [233]. Ainsi, si les études s'appuyant sur des EDS sont à ce jour minoritaires, l'utilisation d'EDS pour la recherche épidémiologique devrait augmenter dans les prochaines années. Cette utilisation s'accompagne néanmoins de nombreux défis.

Défis et leviers d'action

Techniques : normalisation, structuration, exhaustivité et interopérabilité

Les EDS s'appuient sur les DPI et les outils contenus dans les systèmes d'information des établissements de santé, dont l'objectif principal est de tracer et faciliter les soins et non de recueillir des données pour la recherche. Aussi, les données disponibles dans les EDS sont hétérogènes, non ou partiellement structurées, majoritairement narratives, et non exhaustives, donc de qualité variable, ce qui soulève plusieurs difficultés.

Tout d'abord, si les données de vie réelle recueillies et structurées dans les EDS sont très riches, la multiplicité de leurs sources et des acteurs les produisant, l'hétérogénéité de leurs formats et du langage utilisé sont des limites à leur utilisation. Par ailleurs, ces limites impactent directement l'interopérabilité des EDS, c'est-à-dire leur capacité à fonctionner avec d'autres bases de données existantes ou futures. En effet, les patients sont pris en charge par des équipes pluridisciplinaires, dans des cadres et avec des outils et des logiciels différents, générant un volume important de données qui s'accumulent avec le temps. L'amélioration de la qualité, la normalisation et la structuration des données lors de leur inclusion dans un EDS sont des étapes complexes nécessitant des ressources humaines, temporelles et techniques importantes. La complexité et les particularités du domaine médical et du sous-langage médical rendent difficile l'optimisation du développement des EDS sans collaboration étroite avec des soignants. Malheureusement, ces collaborations sont freinées par la diversité des compétences en informatique et

en santé nécessaires à leur bon déroulement, et à la situation de tension croissante du système de santé rendant difficile les initiatives de formation et de mise à disposition de soignants dans ce but.

Par ailleurs, le caractère majoritairement narratif des données induit une perte d'informations importante. Le développement de programmes de TAL est en plein essor et vise à optimiser le potentiel de réutilisation de ces données narratives. Cependant, comme nous l'avons montré à travers un travail exploratoire, le développement de programmes de TAL est complexe, long, et nécessite un travail pluridisciplinaire. Surtout, un programme de TAL répond à une question spécifique dans un contexte précis. La transposition d'un programme de TAL dans un autre contexte nécessite obligatoirement une nouvelle étape d'enrichissement de la terminologie sur laquelle le programme s'appuie. Enfin, la qualité des extractions effectuée par un programme de TAL dépend intrinsèquement de la qualité du matériel qu'on lui apporte, donc de l'exhaustivité de l'information et de l'homogénéité du langage utilisé dans les comptes rendus. Si le TAL peut être utile dans des projets ponctuels, il ne semble pas à ce jour suffisant à lui seul pour pouvoir réutiliser les données narratives des comptes-rendus médicaux pour la recherche, et son potentiel sera amélioré avec l'optimisation du contenu des comptes-rendus.

La non-exhaustivité des données soulève également la problématique des données manquantes. En effet, les données notifiées dans les comptes-rendus médicaux dépendent du choix des soignants sur les données d'intérêt ou non, et de la transposition éventuelle (automatique ou non) de données issues de divers logiciels ou sources. Hors l'absence d'information dans le domaine médical peut à la fois signifier la normalité du paramètre évalué, ou l'absence d'information sur la non-normalité. A titre d'exemple, certains examens biologiques ou physiologiques peuvent ne pas être notés par les soignants dans les comptes-rendus des patients par oubli, ou car ils ne sont pas jugés d'intérêt, ou car ils sont normaux. Dans les travaux présentés dans ce manuscrit, l'absence de code CIM-10 d'infection bactérienne autre que la bactériémie a été interprétée comme l'absence de site primaire de la bactériémie, alors qu'il pourrait s'agir d'informations manquantes.

Plusieurs leviers d'actions existent pour améliorer la qualité des données et donc leur interopérabilité et leur potentiel de réutilisation et de partage.

Tout d'abord, les outils sur lesquels reposent les systèmes d'information des hôpitaux devraient être conçus de manière pluridisciplinaire avec des soignants, des chercheurs et des experts en informatique et en base de données. Les objectifs de ces outils devraient être centrés non seulement autour de la traçabilité de l'information médicale mais aussi de la mise en commun des données pour la recherche et l'optimisation des soins. Ces outils doivent également tenir compte de la situation de tension que connaît le système hospitalier, donc des contraintes temporelles auxquelles font face les soignants quand ils rentrent les informations des patients. Idéalement, il faudrait trouver un juste équilibre entre un outil où les données seraient rentrées de manière structurée et exhaustive, ce qui rendrait l'outil

inutilisable en pratique clinique, et un format narratif, qui restreindrait l'utilisation de l'outil à des fins de stockage des données. Des fonctionnalités pourraient permettre cela, comme la possibilité de transposer et structurer automatiquement certains paramètres cliniques ou paracliniques dans les comptes rendus médicaux et les dossiers des patients, ou la mise en place de correcteurs automatiques qui pourraient améliorer le format et l'exhaustivité des données tout en facilitant le remplissage des données par les soignants.

Ensuite, la collaboration de soignants et de spécialistes en données et en informatique, semble nécessaire à toutes les étapes de la conception et du développement des EDS, depuis la définition de la stratégie de remontée des données prioritaires dans les EDS, à leur structuration et mise à disposition des utilisateurs. Cette collaboration nécessite de former spécifiquement des soignants à l'informatique et aux bases de données, et inversement des experts en base de données et en informatique aux particularités du soin. Afin de maximiser l'interopérabilité des données des EDS, ceux-ci devraient contenir dès leur conception les données nécessaires à leur chaînage avec d'autres bases de données, comme les bases médico-administratives.

Par ailleurs, que cela soit pour le développement des systèmes d'information hospitaliers, de leurs outils, ou des EDS, il est indispensable de s'approcher au maximum de formats communs de structuration, d'organisation et de stockage des données, et d'utiliser des nomenclatures standardisées internationales. Par exemple, les données de l'EDS de l'AP-HP étaient organisées sous un format i2b2, développé il y a une dizaine d'années aux Etats-Unis et utilisé par environ 200 sites dans le Monde [234]. Depuis l'épidémie due à la covid-19, le format des données a transitionné vers le modèle commun de données OMOP-CDM (Observational medical outcomes partnership - Common Data Model), qui a pour objectif l'interopérabilité entre les différentes bases de données en santé, qu'elles soient cliniques ou médico-administratives, à une échelle internationale [235]. Cette transition vient s'ajouter à d'autres travaux d'alignement sur les standards internationaux des données de l'EDS de l'AP-HP (alignement sur les terminologies de référence LOINC³⁰, CIM, etc.) [236]. Pour améliorer la standardisation des données et le potentiel des EDS pour la recherche, une piste pourrait être la mise en place de rencontres dédiées entre les différents concepteurs et utilisateurs de systèmes d'information hospitaliers et d'EDS à l'échelle française, voire européenne. Ces rencontres permettraient de donner de la visibilité aux travaux réalisés grâce aux données des EDS et donc montrer le potentiel de ces données, mais surtout de partager les réflexions, outils et décisions prises pour organiser et structurer les données, de développer des collaborations, et, plus globalement de partager des connaissances et des compétences à ce sujet. Par ailleurs, chaque hôpital avec un EDS devrait être incité à organiser régulièrement des rencontres à son échelle entre les producteurs et utilisateurs de données, afin d'optimiser la recherche sur les données de

³⁰ Logical Observation Identifiers Names & Codes. Il s'agit d'une terminologie de référence internationale.

EDS, et nourrir la réflexion et les échanges interdisciplinaires autour des données de santé. Enfin, la mise en place d'une gouvernance centrale permettant de guider et accompagner les futurs constructeurs d'EDS, telle que l'appel à projet lancé en France pour accompagner la constitution d'EDS avec la coordination du HDH, pourrait également améliorer l'interopérabilité des données et la mise en place de liens réels entre concepteurs et utilisateurs des EDS.

Légaux et éthiques

Si la mise à disposition de grands volumes et d'une grande diversité de données de santé est une avancée majeure pour la recherche et l'optimisation des soins, elle s'accompagne de questionnements légaux et éthiques. En effet, les données de santé sont des données à caractère personnel, considérées comme sensibles. La réutilisation de ces données pour la recherche, et la promotion des données ouvertes et du partage des données, ne peuvent faire l'économie du respect du droit des personnes en termes de droit à l'information, de confidentialité, et de sécurité. Les réglementations à ce sujet sont existantes ou non, et variables, selon les pays. En France, les données de santé font l'objet d'une protection particulière par les textes (règlement général européen sur la protection des données (RGPD), loi Informatique et Libertés, code de la santé publique, etc.) [237]. Notamment, la CNIL a adopté en octobre 2021 un référentiel relatif aux traitements de données à caractère personnel mis en œuvre à des fins de création d'entrepôts de données de santé dont la constitution se fonde sur l'exercice d'une mission d'intérêt public, au sens de l'article 6-1-e du RGPD [238,239]. Ce référentiel définit les données à caractère personnel pouvant être incluses dans un entrepôt, et les exigences en matière de droit à l'information, de durée de conservation des données, d'information des patients et des professionnels, de droits des personnes, de sécurité, de recours à la sous-traitance, de transfert des données en dehors de l'Union européenne, et d'analyse d'impact sur la protection des données [238]. De manière plus générale, une gouvernance poussant vers l'implémentation de cadres légaux pour les EDS à l'échelle européenne voire internationale est nécessaire, pour assurer la protection des données des individus tout en maximisant le potentiel de partage et d'ouverture des données à des fins de recherche et d'optimisation des soins. Une attention particulière devrait être portée à créer des cadres réglementaires qui facilitent le processus et les temporalités d'accès aux données pour les chercheurs, sans pour autant nuire aux exigences de sécurité et de confidentialité en rapport avec les données.



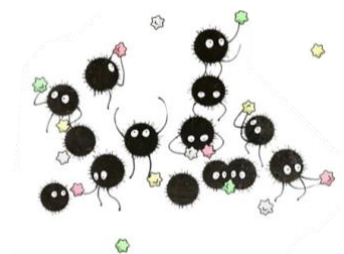
Chainage avec d'autres types de bases de données

Les données des EDS sont par définition hospitalières, ce qui peut constituer une limite importante pour des projets de recherche nécessitant l'accès à des données communautaires. Cette limite a été illustrée dans ce manuscrit à travers l'exemple de l'étude de l'antibiorésistance. Les EDS contiennent des données détaillées cliniques et microbiologiques pour des patients hospitalisés avec des infections bactériennes

résistantes ou non aux antibiotiques. Cependant, l'exposition médicamenteuse et aux soins en dehors de l'hospitalisation n'est pas connue, ce qui empêche d'étudier le parcours de soin des patients avant et après leur hospitalisation et de différencier précisément les infections communautaires vraies de celles communautaires associées aux soins. Par ailleurs, l'évaluation de l'impact de divers facteurs d'exposition médicamenteuse sur l'antibiorésistance n'est pas possible. Dans ce manuscrit, nous avons également présenté les avantages et limites du SNDS, une base de données nationale, centralisée, et fournissant de façon exhaustive des données sociodémographiques, médicales, socioéconomiques et toutes les consommations de soins en ville et à l'hôpital pour l'ensemble de la population en France, à l'exclusion des résultats biologiques, des données physiologiques, d'examen physique, radiologiques, et des traitements reçus dans les établissements de santé (sauf ceux facturés « en sus » des forfaits hospitaliers). Aussi, il apparaît que les EDS et le SNDS sont des sources de données très riches mais avec des limites, qui pourraient être atténuées par leur mise en commun.

Des exemples existent à l'étranger, comme la Veterans Affairs Corporate Data Warehouse qui regroupe les données nationales des DPI des vétérans, provenant des pharmacies, des laboratoires, et plus largement de toute structure de soin ambulatoire ou hospitalière dédiée aux vétérans [68,240,241]. Cette base rassemble 60 domaines de données (statut vital, données démographiques, résultats de laboratoire, commandes de médicaments, administration de médicaments par code-barres, signes vitaux, images de radiologie, etc.) sélectionnés par des experts cliniques et des responsables opérationnels. Cette base couvre ainsi de manière exhaustive les 6 millions de vétérans et contient plus de 10 millions d'hospitalisations, 2,7 milliards de signes vitaux et 6,6 milliards de résultats de laboratoire [68,240,241]. Elle a mené à de très nombreuses publications. Un autre exemple est la base DACOBAN qui chaîne les données hospitalières de microbiologie de trois hôpitaux danois avec le registre national danois de patients, et le système danois d'enregistrement civil, permettant d'avoir des données microbiologiques, le détail des comorbidités et du statut vital des patients, et pour environ 20% des épisodes de bactériémie des données cliniques [121]. La base est continuellement mise à jour, couvre 32% de la population danoise, et fait partie d'un réseau international de registres de bactériémies [121]. Elle a déjà permis la réalisation de multiples travaux de recherche [110,242–248]. Plusieurs projets de chaînage de données d'EDS avec le SNDS sont en cours en France, dont le projet BactHub qui a été présenté précédemment et qui fait toujours l'objet d'une demande d'agrément à la CNIL. Malgré le caractère notablement novateur de tels projets, ils sont aujourd'hui particulièrement difficiles à mettre en place pour les chercheurs en France, et nécessitent parfois plusieurs années avant d'accéder à des données dont l'originalité sera au moment de l'accès périmée. L'étude de cas autour de projets de recherche en cours en France, nécessitant le chaînage d'EDS et du SNDS, devrait être réalisée et discutée avec tous les acteurs impliqués, afin de faire émerger des points d'action pour faciliter et fluidifier les processus réglementaires, et l'accès aux données. Plus que les défis techniques, les défis humains et

réglementaires liés au partage et à la réutilisation des données semblent être de vrais freins aujourd’hui à l’innovation dans ce domaine en France. Il est plus que nécessaire de créer des ponts entre les différents acteurs de l’écosystème des données de santé (patients, soignants, producteurs, utilisateurs des données, acteurs institutionnels, juridiques, industriels et politiques) afin de s’aligner sur des objectifs communs, d’adapter le cadre réglementaire, et donc de tirer le maximum de la richesse des bases de données de santé françaises.



10 CONCLUSION GÉNÉRALE



Cette thèse met en avant la richesse des données cliniques et microbiologiques des entrepôts de données de santé, et leur potentiel pour l'étude des infections bactériennes hospitalisées, résistantes aux antibiotiques. Elle apporte des données récentes sur la morbidité et la mortalité de plus de 30 000 épisodes de bactériémie, et met en exergue les impacts différentiels des bactéries et des résistances aux antibiotiques sur une issue significative, la récurrence infectieuse. Elle discute à travers un exemple l'apport des techniques de traitement automatisé du langage naturel pour la réutilisation des données narratives des comptes-rendus hospitaliers. Enfin, elle souligne les défis liés à l'étude des infections communautaires vraies, et le potentiel des entrepôts de données de santé pour y répondre en cas de chaînage avec les données du SNDS. Il s'agit d'un horizon important de recherche pour les prochaines années, au vu de la diffusion de certaines résistances aux antibiotiques dans la population générale.

11 RÉFÉRENCES

1. Frérot M, Lefebvre A, Aho S, Callier P, Astruc K, Aho Glélé LS. What is epidemiology? Changing definitions of epidemiology 1978-2017. Mason PH, éditeur. PLOS ONE. 10 déc 2018;13(12):e0208442.
2. Djokic T, Van Kranendonk MJ, Campbell KA, Walter MR, Ward CR. Earliest signs of life on land preserved in ca. 3.5 Ga hot spring deposits. Nat Commun. 8 août 2017;8(1):15263.
3. Mills DB, Boyle RA, Daines SJ, Sperling EA, Pisani D, Donoghue PCJ, et al. Eukaryogenesis and oxygen in Earth history. Nat Ecol Evol. mai 2022;6(5):520-32.
4. Krissansen-Totton J, Arney GN, Catling DC. Constraining the climate and ocean pH of the early Earth with a geological carbon cycle model. Proc Natl Acad Sci. 17 avr 2018;115(16):4105-10.
5. Comment, au sortir des océans, la vie a fleuri sur la terre. Steve Parker. The Conversation. Accessible sur le lien : <https://theconversation.com/comment-au-sortir-des-océans-la-vie-a-fleuri-sur-la-terre-102595>.
6. Les origines inattendues de la photosynthèse. Jean-Pierre Jacquot. The Conversation. Accessible sur le lien : <https://theconversation.com/les-origines-inattendues-de-la-photosynthese-62310>.
7. Microbiology by numbers. Nat Rev Microbiol. sept 2011;9(9):628-628.
8. Casadevall A, Pirofski L. Host-Pathogen Interactions: Redefining the Basic Concepts of Virulence and Pathogenicity. Fischetti VA, éditeur. Infect Immun. août 1999;67(8):3703-13.
9. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLOS Biol. 19 août 2016;14(8):e1002533.
10. Lee LH, Wong SH, Chin SF, Singh V, Ab Mutalib NS. Editorial: Human Microbiome: Symbiosis to Pathogenesis. Front Microbiol. 17 févr 2021;12:605783.
11. Brestoff JR, Artis D. Commensal bacteria at the interface of host metabolism and the immune system. Nat Immunol. juill 2013;14(7):676-84.
12. Pirofski L, Casadevall A. The meaning of microbial exposure, infection, colonisation, and disease in clinical practice. Lancet Infect Dis. oct 2002;2(10):628-35.
13. Braz VS, Melchior K, Moreira CG. Escherichia coli as a Multifaceted Pathogenic and Versatile Bacterium. Front Cell Infect Microbiol. 21 déc 2020;10:548492.
14. Jauréguy F. Déterminants cliniques et bactériens au cours des infections extra-intestinales dues à *Escherichia coli*. médecine/sciences. mars 2009;25(3):221-3.
15. Leimbach A, Hacker J, Dobrindt U. E. coli as an All-Rounder: The Thin Line Between Commensalism and Pathogenicity. In: Dobrindt U, Hacker JH, Svanborg C, éditeurs. Between Pathogenicity and Commensalism [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013 [cité 9 mars 2022]. p. 3-32. (Current Topics in Microbiology and Immunology; vol. 358). Disponible sur: http://link.springer.com/10.1007/82_2012_303
16. Adapté du schéma disponible sur le lien : <https://souslemicroscope.com/comment-nommer-une-bacterie/>. Consulté le 16/03/22.
17. Levin BR, Antia R. Why We Don't Get Sick: The Within-Host Population Dynamics of Bacterial Infections. Science. 11 mai 2001;292(5519):1112-5.
18. Données ECDC 2019 pour la France. Accessibles sur le lien : <https://www.ecdc.europa.eu/en/antimicrobial-consumption/database/country-overview>, consulté le 11/03/22.
19. CMIT. Chapitres « Antibiothérapie: principes généraux » et « β -lactamines ». E.PILLY 27eme édition: ALINEA Plus Ed; 2020: pp 26-32 et 37-45. In.
20. Principaux modes d'actions des antibiotiques. Adapté à partir de la figure disponible sur le lien : <http://www.123bio.net/cours/antibio/modedaction.html>. Consulté le 16/03/22.
21. Klein EY, Van Boeckel TP, Martinez EM, Pant S, Gandra S, Levin SA, et al. Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. Proc Natl Acad Sci [Internet]. 10 avr 2018 [cité 11 mars 2022];115(15). Disponible sur: <https://pnas.org/doi/full/10.1073/pnas.1717295115>

22. European Centre for Disease Prevention and Control. Antimicrobial consumption in the EU/EEA – Annual Epidemiological Report 2019. Stockholm: ECDC; 2020.
23. Cavalié P, Le Vu S, Jezewski-Serra D, Maugat S, Berger-Carbonne A. Consommation d'antibiotiques en secteur de ville en France de 2009 à 2019. Synthèse des indicateurs mis en ligne sur Géodes. Saint-Maurice: Sante publique France; novembre 2020, 7 p.
24. Données GEODES 2019. Accessibles sur le lien : <https://geodes.santepubliquefrance.fr>, consulté le 11/03/22.
25. Consommation d'antibiotiques et antibiorésistance en France en 2019. Santé Publique France. Accessible sur le lien : <https://www.santepubliquefrance.fr/les-actualites/2020/consommation-d-antibiotiques-et-antibioresistance-en-france-en-2019>.
26. Plachouras D, Kärki T, Hansen S, Hopkins S, Lyytikäinen O, Moro ML, et al. Antimicrobial use in European acute care hospitals: results from the second point prevalence survey (PPS) of healthcare-associated infections and antimicrobial use, 2016 to 2017. *Eurosurveillance* [Internet]. 15 nov 2018 [cité 11 mars 2022];23(46). Disponible sur: <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.23.46.1800393>
27. World Health Organization. Global action plan on antimicrobial resistance [Internet]. 2015. Disponible sur: <https://www.who.int/antimicrobial-resistance/global-action-plan/en/>
28. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*. févr 2022;399(10325):629-55.
29. Opatowski M, Tuppin P, Cosker K, Touat M, De Lagasnerie G, Guillemot D, et al. Hospitalisations with infections related to antimicrobial-resistant bacteria from the French nationwide hospital discharge database, 2016. *Epidemiol Infect*. 2019;147:e144.
30. Touat M, Opatowski M, Brun-Buisson C, Cosker K, Guillemot D, Salomon J, et al. A Payer Perspective of the Hospital Inpatient Additional Care Costs of Antimicrobial Resistance in France: A Matched Case-Control Study. *Appl Health Econ Health Policy*. juin 2019;17(3):381-9.
31. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis*. janv 2019;19(1):56-66.
32. Bennett PM. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria: Plasmid-encoded antibiotic resistance. *Br J Pharmacol*. mars 2008;153(S1):S347-57.
33. Bactéries multirésistantes dans les hôpitaux français: des premiers indicateurs au Réseau d'alerte d'investigation et de surveillance des infections nosocomiales (Raisin). Jarlier Vincent. *Bulletin Epidémiologique Hebdomadaire*, 2004, n°. 32-33, p. 148-51. Accessible sur le lien: <https://www.santepubliquefrance.fr/maladies-et-traumatismes/infections-associees-aux-soins-et-resistance-aux-antibiotiques/infections-associees-aux-soins/documents/article/bacteries-multiresistantes-dans-les-hopitaux-francais-des-premiers-indicateurs-au-reseau-d-alerte-d-investigation-et-de-surveillance-des-infectio>.
34. Des réseaux de surveillance nationaux. Santé Publique France. Accessible sur le lien: <https://www.santepubliquefrance.fr/maladies-et-traumatismes/infections-associees-aux-soins-et-resistance-aux-antibiotiques/resistance-aux-antibiotiques/articles/des-reseaux-de-surveillance-nationaux>. Consulté le 15/03/22.
35. CMIT. Chapitres « Infections associées aux soins - hygiène hospitalière: généralités ». E.PILLY 27eme édition: ALINEA Plus Ed; 2020: pp 618-622. In.
36. Principaux mécanismes de résistance aux antibiotiques. Adapté à partir de la figure disponible sur le lien: https://www.reactgroup.org/wp-content/uploads/2016/09/Resistance-mechanisms_Erik-Gullberg.png. Consulté le 16/03/22.
37. Antimicrobial resistance in the EU/EEA (EARS-Net) - Annual Epidemiological Report for 2019. Available from: <https://www.ecdc.europa.eu/sites/default/files/documents/surveillance-antimicrobial-resistance-Europe-2019.pdf>.
38. Bell BG, Schellevis F, Stobberingh E, Goossens H, Pringle M. A systematic review and meta-

- analysis of the effects of antibiotic consumption on antibiotic resistance. *BMC Infect Dis* [Internet]. déc 2014 [cité 27 janv 2019];14(1). Disponible sur: <https://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-14-13>
39. Costelloe C, Metcalfe C, Lovering A, Mant D, Hay AD. Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: systematic review and meta-analysis. *BMJ*. 2 juin 2010;340(may18 2):c2096-c2096.
 40. Bronzwaer S, Lönnroth A, Haigh R. The European Community Strategy Against Antimicrobial Resistance. *Eurosurveillance*. 1 janv 2004;9(1):30-4.
 41. for the World Healthcare-Associated Infections Resistance Forum participants, Harbarth S, Balkhy HH, Goossens H, Jarlier V, Kluytmans J, et al. Antimicrobial resistance: one world, one fight! *Antimicrob Resist Infect Control*. déc 2015;4(1):49.
 42. Stratégie nationale 2022-2025 de Prévention des Infections et de l'Antibiorésistance en santé humaine, Ministère des Solidarités et de la Santé, France. Accessible sur le lien: https://solidarites-sante.gouv.fr/IMG/pdf/strategie_nationale_2022-2025_prevention_des_infections_et_de_l_antibioresistance.pdf, consulté le 11/03/22.
 43. Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé L'exemple du médicament. Bernard Bégaud, Dominique Polton, Franck von Lenep. Rapport réalisé à la demande de Madame la Ministre de la santé Marisol Touraine. Mai 2017. Accessible sur le lien: https://solidarites-sante.gouv.fr/IMG/pdf/rapport_donnees_de_vie_reelle_medicaments_mai_2017vf.pdf.
 44. Valles J, Leon C, Spanish Collaborative Group for Infections in Intensive Care Units of Sociedad Espanola de Medicina Intensiva y Unidades Coronarias (SEMIUC). Nosocomial Bacteremia in Critically Ill Patients: A Multicenter Study Evaluating Epidemiology and Prognosis. *Clin Infect Dis*. 1 mars 1997;24(3):387-95.
 45. Chang FY, Peacock JE, Musher DM, Triplett P, MacDonald BB, Mylotte JM, et al. Staphylococcus aureus Bacteremia: Recurrence and the Impact of Antibiotic Treatment in a Prospective Multicenter Study. *Medicine (Baltimore)*. sept 2003;82(5):333-9.
 46. Magret M, Lisboa T, Martin-Loeches I, Máñez R, Nauwynck M, Wrigge H, et al. Bacteremia is an independent risk factor for mortality in nosocomial pneumonia: a prospective and observational multicenter study. *Crit Care*. 2011;15(1):R62.
 47. Loi 2012-300 du 5 mars 2012 relatives aux recherches impliquant la personne humaine. Accessible sur le lien: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000025441587/>.
 48. Micek ST, Wunderink RG, Kollef MH, Chen C, Rello J, Chastre J, et al. An international multicenter retrospective study of Pseudomonas aeruginosa nosocomial pneumonia: impact of multidrug resistance. *Crit Care*. déc 2015;19(1):219.
 49. Letaief A, Boughzala E, Kaabia N, Ernez S, Abid F, Chaabane TB, et al. Epidemiology of infective endocarditis in Tunisia: a 10-year multicenter retrospective study. *Int J Infect Dis*. sept 2007;11(5):430-3.
 50. Ramirez J, Aliberti S, Mirsaeidi M, Peyrani P, Filardo G, Amir A, et al. Acute Myocardial Infarction in Hospitalized Patients with Community-Acquired Pneumonia. *Clin Infect Dis*. 15 juill 2008;47(2):182-7.
 51. Mission Spares 2018. Bactéries multirésistantes en établissements de santé en 2018. Mission nationale Spares, novembre 2019. Données 2018 du réseau BMR-Raisin. Saint-Maurice: Sante publique France; septembre 2019. 6 p. Disponible a partir de l'URL: <https://www.santepubliquefrance.fr/>.
 52. Mission Spares 2019. Surveillance de la consommation des antibiotiques et des résistances bactériennes en établissement de santé. Mission Spares, résultats 2019. Saint-Maurice: Santé publique France, 2021. 80 p. Disponible à partir de l'URL: www.santepubliquefrance.fr.
 53. Mission Spares 2020. Surveillance de l'antibiorésistance en établissements de santé. Mission Spares. Résultats 2020. Saint-Maurice: Santé publique France, 2022. 107 p. Disponible à partir de l'URL: www.santepubliquefrance.fr.
 54. Mission Primo 2018. Surveillance de la résistance bactérienne aux antibiotiques en soins de ville et en établissements pour personnes âgées dépendantes. Réseau Primo: résultats 2018. Saint-Maurice: Santé publique France, 2019. 66 p. Accessible sur le lien: www.santepubliquefrance.fr.
 55. Mission Primo 2019. Surveillance de la résistance bactérienne aux antibiotiques en soins de ville

- et en établissements pour personnes âgées dépendantes. Réseau Primo : résultats 2019. Saint-Maurice : Santé publique France, 2021. 72 p. Accessible sur le lien : www.santepubliquefrance.fr.
56. Mission Primo 2020. Surveillance de la résistance bactérienne aux antibiotiques en soins de ville et en établissements pour personnes âgées dépendantes. Réseau Primo : résultats 2020. Saint-Maurice : Santé publique France, 2022. 88 p. Accessible sur le lien : <http://www.santepubliquefrance.fr>.
57. Outil MedQual-Ville. Présentation et résultats accessibles sur le lien : <https://medqualville.antibioreistance.fr/>.
58. Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev D'Épidémiologie Santé Publique*. oct 2017;65:S149-67.
59. Composantes du SNDS. Accessible sur le lien : <https://www.snds.gouv.fr/SNDS/Composantes-du-SNDS>. Consulté le 20/05/2022.
60. Médicaments de la liste en sus. Documentation du SNDS. Accessible sur le lien : https://documentation-snds.health-data-hub.fr/fiches/medicaments_de_la_liste_en_sus.html#presentation.
61. Khalaf Hamoud A, Salah Hashim A, Akeel Awadh W. CLINICAL DATA WAREHOUSE A REVIEW. *Iraqi J Comput Inform [Internet]*. 31 déc 2018 [cité 6 déc 2021];44(2). Disponible sur : <http://www.uoitc.edu.iq/ijci1/vol44issue2/4-CLINICALDATAWAREHOUSEAREVIEW.pdf>
62. Lovis C. Systèmes d'informations cliniques: au coeur des informatiques hospitalières. In: *Risques, Technologies de l'Information pour les Pratiques Médicales [Internet]*. Paris: Springer Paris; 2009 [cité 12 mars 2022]. p. 35-44. (Informatique et Santé; vol. 17). Disponible sur : http://link.springer.com/10.1007/978-2-287-99305-3_4
63. Zapletal E, Rodon N, Grabar N, Degoulet P. Methodology of integration of a clinical data warehouse with a clinical information system: the HEGP case. *Stud Health Technol Inform*. 2010;160(Pt 1):193-7.
64. Atlas des SIH 2020. Agence technique de l'informatisation sur l'hospitalisation, et la direction générale de l'offre de soins du Ministère des solidarités et de la santé. Novembre 2021. Accessible sur le lien : <https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/e-sante/sih/article/atlas-des-systemes-d-information-hospitaliers>.
65. Gillum RF. From Papyrus to the Electronic Tablet: A Brief History of the Clinical Medical Record with Lessons for the Digital Age. *Am J Med*. oct 2013;126(10):853-7.
66. Evans RS. Electronic Health Records: Then, Now, and in the Future. *Yearb Med Inform*. août 2016;25(S 01):S48-61.
67. Inmon WH. *Building the Data Warehouse*. John Wiley, 1992.
68. Wang XQ, Vincent BM, Wiitala WL, Luginbill KA, Viglianti EM, Prescott HC, et al. Veterans Affairs patient database (VAPD 2014–2017): building nationwide granular data for clinical discovery. *BMC Med Res Methodol*. déc 2019;19(1):94.
69. Jannot AS, Zapletal E, Avillach P, Mamzer MF, Burgun A, Degoulet P. The Georges Pompidou University Hospital Clinical Data Warehouse: A 8-years follow-up experience. *Int J Med Inf*. juin 2017;102:21-8.
70. de Mul M, Alons P, van der Velde P, Konings I, Bakker J, Hazelzet J. Development of a clinical data warehouse from an intensive care clinical information system. *Comput Methods Programs Biomed*. janv 2012;105(1):22-30.
71. McDanel JS, Roghmann MC, Perencevich EN, Ohl ME, Goto M, Livorsi DJ, et al. Comparative Effectiveness of Cefazolin Versus Nafcillin or Oxacillin for Treatment of Methicillin-Susceptible *Staphylococcus aureus* Infections Complicated by Bacteremia: A Nationwide Cohort Study. *Clin Infect Dis*. 1 juill 2017;65(1):100-6.
72. Albertson J, McDanel JS, Carnahan R, Chrischilles E, Perencevich EN, Goto M, et al. Determination of Risk Factors for Recurrent Methicillin-Resistant *Staphylococcus aureus* Bacteremia in a Veterans Affairs Healthcare System Population. *Infect Control Hosp Epidemiol*. mai 2015;36(5):543-9.
73. Baum A, Schwartz MD. Admissions to Veterans Affairs Hospitals for Emergency Conditions

During the COVID-19 Pandemic. *JAMA*. 7 juill 2020;324(1):96.

74. Hoertel N, Sánchez-Rico M, Vernet R, Beeker N, Jannot AS, Neuraz A, et al. Association between antidepressant use and reduced risk of intubation or death in hospitalized patients with COVID-19: results from an observational study. *Mol Psychiatry*. sept 2021;26(9):5199-212.
75. Kramer JR, Puenpatom A, Erickson KF, Cao Y, Smith D, El-Serag HB, et al. Real-world effectiveness of elbasvir/grazoprevir in HCV-infected patients in the US veterans affairs healthcare system. *J Viral Hepat*. nov 2018;25(11):1270-9.
76. Daniel C, Serre P, Orlova N, Bréant S, Paris N, Griffon N. Initializing a hospital-wide data quality program. The AP-HP experience. *Comput Methods Programs Biomed*. nov 2019;181:104804.
77. Lien: <https://www.cnil.fr/fr/la-cnil-adopte-un-referentiel-sur-les-entrepots-de-donnees-de-sante>. Consulté le 15/03/22.
78. Sogaard M, Norgaard M, Dethlefsen C, Schonheyder HC. Temporal Changes in the Incidence and 30-Day Mortality associated with Bacteremia in Hospitalized Patients from 1992 through 2006: A Population-based Cohort Study. *Clin Infect Dis*. 1 janv 2011;52(1):61-9.
79. Nielsen SL, Lassen AT, Kolmos HJ, Jensen TG, Gradel KO, Pedersen C. The daily risk of bacteremia during hospitalization and associated 30-day mortality evaluated in relation to the traditional classification of bacteremia. *Am J Infect Control*. févr 2016;44(2):167-72.
80. Weinstein MP, Towns ML, Quartey SM, Mirrett S, Reimer LG, Parmigiani G, et al. The Clinical Significance of Positive Blood Cultures in the 1990s: A Prospective Comprehensive Evaluation of the Microbiology, Epidemiology, and Outcome of Bacteremia and Fungemia in Adults. *Clin Infect Dis*. avr 1997;24(4):584-602.
81. Bates DW. Contaminant Blood Cultures and Resource Utilization: The True Consequences of False-Positive Results. *JAMA*. 16 janv 1991;265(3):365.
82. Richter SS, Beekmann SE, Croco JL, Diekema DJ, Koontz FP, Pfaller MA, et al. Minimizing the Workup of Blood Culture Contaminants: Implementation and Evaluation of a Laboratory-Based Algorithm. *J Clin Microbiol*. juill 2002;40(7):2437-44.
83. Weinstein MP. Blood Culture Contamination: Persisting Problems and Partial Progress. *J Clin Microbiol*. juin 2003;41(6):2275-8.
84. Définition des infections associées aux soins. Comité technique des infections nosocomiales et des infections liées aux soins. Direction générale de la santé, et Direction de l'hospitalisation et de l'organisation des soins, Ministère de la santé, de la jeunesse et des sports. Mai 2007. Accessible sur le lien: https://solidarites-sante.gouv.fr/IMG/pdf/rapport_vcourte.pdf. Consulté le 16/03/22.
85. Liste des organismes du Réseau National de Sécurité en Santé (National Healthcare Safety Network) des Etats-Unis. Accessible sur le lien: <http://www.cdc.gov/nhsn/xls/master-organism-com-commensals-lists.xlsx>. Consulté le 16/03/22.
86. Bloodstream Infection Event (Central Line-Associated Bloodstream Infection and Non-central Line Associated Bloodstream Infection). Module des infections associées aux dispositifs médicaux. Réseau National de Sécurité en Santé des Etats-Unis (National Healthcare Safety Network), CDC. Accessible sur le lien: https://www.cdc.gov/nhsn/pdfs/pscmanual/4psc_clabscurrent.pdf. Consulté le 16/03/22.
87. Friedman ND, Kaye KS, Stout JE, McGarry SA, Trivette SL, Briggs JP, et al. Health care--associated bloodstream infections in adults: a reason to change the accepted definition of community-acquired infections. *Ann Intern Med*. 19 nov 2002;137(10):791-7.
88. Abbara S, Guillemot D, Brun-Buisson C, Watier L. From Pathophysiological Hypotheses to Case-Control Study Design: Resistance from Antibiotic Exposure in Community-Onset Infections. *Antibiotics*. 4 févr 2022;11(2):201.
89. Brun-Buisson C, Doyon F, Carlet J. Bacteremia and severe sepsis in adults: a multicenter prospective survey in ICUs and wards of 24 hospitals. French Bacteremia-Sepsis Study Group. *Am J Respir Crit Care Med*. sept 1996;154(3):617-24.
90. Ortega M, Almela M, Martinez JA, Marco F, Soriano A, López J, et al. Epidemiology and outcome of primary community-acquired bacteremia in adult patients. *Eur J Clin Microbiol Infect Dis*. 28 juin 2007;26(7):453-7.

91. L'AP-HP met en place un entrepôt de données de santé pour faciliter la recherche. Tic santé. Accessible sur le lien : <https://www.ticsante.com/story?ID=3191>.
92. Méthodologie de référence MR004. CNIL. Accessible sur le lien : <https://www.cnil.fr/fr/declaration/mr-004-recherches-nimpliquant-pas-la-personne-humaine-etudes-et-evaluations-dans-le>. Consulté le 16/03/22.
93. Présentation ORBIS à la commission médicale d'établissement (CME) de l'AP-HP du 09/11/21. Accessible sur le lien : http://cme.aphp.fr/sites/default/files/CMEDoc/cme9novembre2021_orbis_v2_0.pdf. Consulté le 16/03/22.
94. Documentation de l'EDS de l'AP-HP. Accessible sur le lien : https://doc.eds.aphp.fr/donnees_dispo. Consulté le 16/03/22.
95. Documentation des données disponibles dans l'EDS de l'AP-HP, relatives au circuit du médicament. Accessible sur le lien : https://doc.eds.aphp.fr/donnees_dispo/circuit-medicament. Consulté le 16/03/22.
96. Rapports du Comité de l'Antibiogramme de la Société Française de Microbiologie. Accessibles sur le lien : <https://www.sfm-microbiologie.org/presentation-de-la-sfm/sections-et-groupes-de-travail/comite-de-lantibiogramme/>.
97. Skull SA, Andrews RM, Byrnes GB, Campbell DA, Nolan TM, Brown GV, et al. ICD-10 codes are a valid tool for identification of pneumonia in hospitalized patients aged ≥ 65 years. *Epidemiol Infect.* févr 2008;136(2):232-40.
98. Jolley RJ, Quan H, Jetté N, Sawka KJ, Diep L, Goliath J, et al. Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data. *BMJ Open.* déc 2015;5(12):e009487.
99. Liu F, Weng C, Yu H. Natural Language Processing, Electronic Health Records, and Clinical Research. In: Richesson RL, Andrews JE, éditeurs. *Clinical Research Informatics* [Internet]. London: Springer London; 2012 [cité 23 mai 2022]. p. 293-310. (Health Informatics). Disponible sur : http://link.springer.com/10.1007/978-1-84882-448-5_16
100. Une petite introduction au Traitement Automatique des Langues Naturelles. François Yvon. Accessible sur le lien : <https://perso.limsi.fr/anne/coursM2R/intro.pdf>.
101. Friedman C, Elhadad N. Natural Language Processing in Health Care and Biomedicine. In: Shortliffe EH, Cimino JJ, éditeurs. *Biomedical Informatics* [Internet]. London: Springer London; 2014 [cité 24 mai 2022]. p. 255-84. Disponible sur : http://link.springer.com/10.1007/978-1-4471-4474-8_8
102. Grishman R, Kittredge R. *Analyzing language in restricted domains: sublanguage description and processing*. New York: Routledge; 1986.
103. Harris Z. *Mathematical structures of language*. New York: Wiley; 1968.
104. Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. brat: a Web-based Tool for NLP-Assisted Text Annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107, Avignon, France. Association for Computational Linguistics. In.
105. Outil brat. Introduction accessible sur le lien : <https://brat.nlplab.org/introduction.html>.
106. The medico-economic department of the Ile-de-France Regional Health Agency (Agence Régionale de Santé). Rapport d'activité 2018 des établissements de santé d'Ile-de-France. Available at: <https://www.iledefrance.ars.sante.fr/media/38323/download>.
107. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol.* 15 mars 2011;173(6):676-82.
108. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data: *Med Care.* nov 2005;43(11):1130-9.
109. Statistique annuelle des établissements de santé. Accessibles sur le lien : <https://www.sae-diffusion.sante.gouv.fr/sae-diffusion/recherche.htm>.
110. for the Danish Collaborative Bacteraemia Network (DACOBAN), Gradel KO, Jensen US, Schønheyder HC, Østergaard C, Knudsen JD, et al. Impact of appropriate empirical antibiotic treatment

on recurrence and mortality in patients with bacteraemia: a population-based cohort study. BMC Infect Dis. déc 2017;17(1):122.

111. Jensen US, Knudsen JD, Wehberg S, Gregson DB, Laupland KB. Risk factors for recurrence and death after bacteraemia: a population-based study. Clin Microbiol Infect. août 2011;17(8):1148-54.

112. Rieg S, Peyerl-Hoffmann G, de With K, Theilacker C, Wagner D, Hübner J, et al. Mortality of *S. aureus* bacteremia and infectious diseases specialist consultation – A study of 521 patients in Germany. J Infect. oct 2009;59(4):232-9.

113. Tansarli GS, Andreatos N, Pliakos EE, Mylonakis E. A Systematic Review and Meta-analysis of Antibiotic Treatment Duration for Bacteremia Due to *Enterobacteriaceae*. Antimicrob Agents Chemother [Internet]. mai 2019 [cité 4 nov 2021];63(5). Disponible sur: <https://journals.asm.org/doi/10.1128/AAC.02495-18>

114. Nielsen SL, Lassen AT, Gradel KO, Jensen TG, Kolmos HJ, Hallas J, et al. Bacteremia is associated with excess long-term mortality: A 12-year population-based cohort study. J Infect. févr 2015;70(2):111-26.

115. Botelho J, Grosso F, Peixe L. Antibiotic resistance in *Pseudomonas aeruginosa* – Mechanisms, epidemiology and evolution. Drug Resist Updat. mai 2019;44:100640.

116. Voor In 't Holt AF, Severin JA, Lesaffre EMEH, Vos MC. A systematic review and meta-analyses show that carbapenem use and medical devices are the leading risk factors for carbapenem-resistant *Pseudomonas aeruginosa*. Antimicrob Agents Chemother. mai 2014;58(5):2626-37.

117. Nielsen SL. The incidence and prognosis of patients with bacteremia. Dan Med J. juill 2015;62(7):B5128.

118. Hoertel N, Sánchez-Rico M, Vernet R, Jannot AS, Neuraz A, Blanco C, et al. Observational study of haloperidol in hospitalized patients with COVID-19. Cheungpasitporn W, éditeur. PLOS ONE. 19 févr 2021;16(2):e0247122.

119. Kramer J, Hartman C, White D, Royse K, Richardson P, Thrift A, et al. Validation of HIV -infected cohort identification using automated clinical data in the Department of Veterans Affairs. HIV Med. 26 mai 2019;hiv.12757.

120. Public Health Council (Haut Conseil de la Santé Publique). Conduite à tenir lors d'épisodes de cas groupés d'infections cutanées suppuratives liées aux souches de SARM Co, 2014. Available at: <https://www.hcsp.fr/explore.cgi/avisrapportsdomaine?clefr=453>.

121. Gradel K, Arpi M, Knudsen J, Schönheyder H, Ostergaard C, Sogaard M. The Danish Collaborative Bacteraemia Network (DACOBAN) database. Clin Epidemiol. sept 2014;301.

122. Laupland KB, Schönheyder HC, Kennedy KJ, Lyytikäinen O, Valiquette L, Galbraith J, et al. Rationale for and protocol of a multi-national population-based bacteremia surveillance collaborative. BMC Res Notes. 2009;2(1):146.

123. Lee H, Yoon EJ, Kim D, Jeong SH, Shin JH, Shin JH, et al. Establishment of the South Korean national antimicrobial resistance surveillance system, Kor-GLASS, in 2016. Eurosurveillance [Internet]. 18 oct 2018 [cité 2 nov 2018];23(42). Disponible sur: <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2018.23.42.1700734>

124. Al-Hasan MN, Eckel-Passow JE, Baddour LM. Recurrent gram-negative bloodstream infection: A 10-year population-based cohort study. J Infect. juill 2010;61(1):28-33.

125. Aurora A, Le TD, Akers KS, Blyth DM, Graybill JC, Clemens MS, et al. Recurrent bacteremia: A 10-year retrospective study in combat-related burn casualties. Burns. mai 2019;45(3):579-88.

126. Jensen US, Knudsen JD, Østergaard C, Gradel KO, Frimodt-Møller N, Schönheyder HC. Recurrent bacteraemia: A 10-year regional population-based study of clinical and microbiological risk factors. J Infect. mars 2010;60(3):191-9.

127. Sawatwong P, Sapchookul P, Whistler T, Gregory CJ, Sangwichian O, Makprasert S, et al. High Burden of Extended-Spectrum β -Lactamase-Producing *Escherichia coli* and *Klebsiella pneumoniae* Bacteremia in Older Adults: A Seven-Year Study in Two Rural Thai Provinces. Am J Trop Med Hyg. 3 avr 2019;100(4):943-51.

128. Wendt C, Messer SA, Hollis RJ, Pfaller MA, Wenzel RP, Herwaldt LA. Recurrent Gram-Negative Bacteremia: Incidence and Clinical Patterns. Clin Infect Dis. mars 1999;28(3):611-7.

129. Huang YT, Liao CH, Teng LJ, Hsu HS, Hsueh PR. Reinfection and relapse of recurrent bacteremia caused by *Klebsiella pneumoniae* in a medical center in Taiwan. *Future Microbiol.* sept 2016;11(9):1157-65.
130. Capdevila JA. Incidence and Risk Factors of Recurrent Episodes of Bacteremia in Adults. *Arch Intern Med.* 28 févr 1994;154(4):411.
131. Miller PJ, Farr BM. Morbidity and Mortality Associated with Multiple Episodes of Nosocomial Bloodstream Infection: A Cohort Study. *Infect Control Hosp Epidemiol.* mai 1989;10(5):216-9.
132. Woudt SHS, de Greeff SC, Schoffelen AF, Vlek ALM, Bonten MJM, Infectious Diseases Surveillance Information System–Antimicrobial Resistance (ISIS-AR) Study Group, et al. Antibiotic Resistance and the Risk of Recurrent Bacteremia. *Clin Infect Dis.* 17 mai 2018;66(11):1651-7.
133. Choi SH, Dagher M, Ruffin F, Park LP, Sharma-Kuinkel BK, Souli M, et al. Risk Factors for Recurrent *Staphylococcus aureus* Bacteremia. *Clin Infect Dis.* 1 juin 2021;72(11):1891-9.
134. Turett GS, Blum S, Telzak EE. Recurrent Pneumococcal Bacteremia: Risk Factors and Outcomes. *Arch Intern Med.* 24 sept 2001;161(17):2141.
135. Patolia S, Abate G, Patel N, Patolia S, Frey S. Risk factors and outcomes for multidrug-resistant Gram-negative bacilli bacteremia. *Ther Adv Infect Dis.* janv 2018;5(1):11-8.
136. Sanz-García M, Fernández-Cruz A, Rodríguez-Crèixems M, Cercenado E, Marin M, Muñoz P, et al. Recurrent *Escherichia coli* Bloodstream Infections: Epidemiology and Risk Factors. *Medicine (Baltimore).* mars 2009;88(2):77-82.
137. Mylotte JM, McDermott C. Recurrent gram-negative bacteremia. *Am J Med.* août 1988;85(2):159-63.
138. Fowler VG, Kong LK, Corey GR, Gottlieb GS, McClelland RS, Sexton DJ, et al. Recurrent *Staphylococcus aureus* Bacteremia: Pulsed-Field Gel Electrophoresis Findings in 29 Patients. *J Infect Dis.* mai 1999;179(5):1157-61.
139. Lee CH, Su LH, Chen FJ, Tang YF, Chien CC, Liu JW. Clinical and microbiologic characteristics of adult patients with recurrent bacteraemia caused by extended-spectrum β -lactamase-producing *Escherichia coli* or *Klebsiella pneumoniae*. *Clin Microbiol Infect.* déc 2015;21(12):1105.e1-1105.e8.
140. Harris PNA, Peri AM, Pelecanos AM, Hughes CM, Paterson DL, Ferguson JK. Risk factors for relapse or persistence of bacteraemia caused by *Enterobacter* spp.: a case–control study. *Antimicrob Resist Infect Control.* déc 2017;6(1):14.
141. Coque TM, Baquero F, Cantón R. Increasing prevalence of ESBL-producing Enterobacteriaceae in Europe. *Eurosurveillance* [Internet]. 20 nov 2008 [cité 14 juill 2022];13(47). Disponible sur: <https://www.eurosurveillance.org/content/10.2807/ese.13.47.19044-en>
142. Gauzit R, Castan B, Bonnet E, Bru JP, Cohen R, Diamantis S, et al. Anti-infectious treatment duration: The SPILF and GPIP French guidelines and recommendations. *Infect Dis Now.* mars 2021;51(2):114-39.
143. Chang D, Sharma L, Dela Cruz CS, Zhang D. Clinical Epidemiology, Risk Factors, and Control Strategies of *Klebsiella pneumoniae* Infection. *Front Microbiol.* 22 déc 2021;12:750662.
144. Scheuerman O, Schechner V, Carmeli Y, Gutiérrez-Gutiérrez B, Calbo E, Almirante B, et al. Comparison of Predictors and Mortality Between Bloodstream Infections Caused by ESBL-Producing *Escherichia coli* and ESBL-Producing *Klebsiella pneumoniae*. *Infect Control Hosp Epidemiol.* juin 2018;39(6):660-7.
145. Burnham JP, Kwon JH, Olsen MA, Babcock HM, Kollef MH. Differences in mortality between infections due to extended-spectrum-beta-lactamase–producing *Klebsiella pneumoniae* and *Escherichia coli*. *Infect Control Hosp Epidemiol.* sept 2018;39(9):1138-9.
146. Ruppé E, Armand-Lefèvre L, Estellat C, Consigny PH, El Mniai A, Boussadia Y, et al. High Rate of Acquisition but Short Duration of Carriage of Multidrug-Resistant Enterobacteriaceae After Travel to the Tropics. *Clin Infect Dis Off Publ Infect Dis Soc Am.* 15 août 2015;61(4):593-600.
147. Titelman E, Hasan CM, Iversen A, Naucleur P, Kais M, Kalin M, et al. Faecal carriage of extended-spectrum β -lactamase-producing Enterobacteriaceae is common 12 months after infection and is related to strain factors. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis.* août 2014;20(8):O508-515.

148. Haverkate MR, Derde LPG, Brun-Buisson C, Bonten MJM, Bootsma MCJ. Duration of colonization with antimicrobial-resistant bacteria after ICU discharge. *Intensive Care Med.* avr 2014;40(4):564-71.
149. Birgand G, Armand-Lefevre L, Lolom I, Ruppe E, Andremont A, Lucet JC. Duration of colonization by extended-spectrum β -lactamase-producing Enterobacteriaceae after hospital discharge. *Am J Infect Control.* mai 2013;41(5):443-7.
150. Pilmis B, Le Monnier A, Zahar JR. Gut Microbiota, Antibiotic Therapy and Antimicrobial Resistance: A Narrative Review. *Microorganisms.* 17 févr 2020;8(2):269.
151. Schechner V, Temkin E, Harbarth S, Carmeli Y, Schwaber MJ. Epidemiological Interpretation of Studies Examining the Effect of Antibiotic Usage on Resistance. *Clin Microbiol Rev.* 1 avr 2013;26(2):289-307.
152. Harris AD, Karchmer TB, Carmeli Y, Samore MH. Methodological Principles of Case-Control Studies That Analyzed Risk Factors for Antibiotic Resistance: A Systematic Review. *Clin Infect Dis.* 1 avr 2001;32(7):1055-61.
153. Harris AD, Carmeli Y, Samore MH, Kaye KS, Perencevich E. Impact of Severity of Illness Bias and Control Group Misclassification Bias in Case-Control Studies of Antimicrobial-Resistant Organisms. *Infect Control Hosp Epidemiol.* 2005;26(4):342-5.
154. Harris AD, Samore MH, Carmeli Y. Control group selection is an important but neglected issue in studies of antibiotic resistance. *Ann Intern Med.* 18 juill 2000;133(2):159.
155. Kaye KS, Harris AD, Samore M, Carmeli Y. The Case-Case-Control Study Design: Addressing the Limitations of Risk Factor Studies for Antimicrobial Resistance. *Infect Control Hosp Epidemiol.* avr 2005;26(4):346-51.
156. Olesen SW, Barnett ML, MacFadden DR, Brownstein JS, Hernández-Díaz S, Lipsitch M, et al. The distribution of antibiotic use and its association with antibiotic resistance. *eLife.* 18 déc 2018;7:e39435.
157. Lipsitch M, Samore MH. Antimicrobial use and antimicrobial resistance: a population perspective. *Emerg Infect Dis.* avr 2002;8(4):347-54.
158. Nicolas-Chanoine MH, Gruson C, Bialek-Davenet S, Bertrand X, Thomas-Jean F, Bert F, et al. 10-Fold increase (2006-11) in the rate of healthy subjects with extended-spectrum β -lactamase-producing *Escherichia coli* faecal carriage in a Parisian check-up centre. *J Antimicrob Chemother.* 1 mars 2013;68(3):562-8.
159. Rodríguez-Baño J, Picón E, Gijón P, Hernández JR, Ruíz M, Peña C, et al. Community-Onset Bacteremia Due to Extended-Spectrum β -Lactamase-Producing *Escherichia coli*: Risk Factors and Prognosis. *Clin Infect Dis.* janv 2010;50(1):40-8.
160. Hsieh CJ, Shen YH, Hwang KP. Clinical Implications, Risk Factors and Mortality Following Community-onset Bacteremia Caused by Extended-spectrum β -lactamase (ESBL) and non-ESBL Producing *Escherichia coli*. *J Microbiol Immunol Infect.* juin 2010;43(3):240-8.
161. Park SH, Choi SM, Lee DG, Kim J, Choi JH, Kim SH, et al. Emergence of Extended-Spectrum β -Lactamase-Producing *Escherichia coli* as a Cause of Community-Onset Bacteremia in South Korea: Risk Factors and Clinical Outcomes. *Microb Drug Resist.* déc 2011;17(4):537-44.
162. Park YS, Bae IK, Kim J, Jeong SH, Hwang S sik, Seo YH, et al. Risk Factors and Molecular Epidemiology of Community-Onset Extended-Spectrum β -Lactamase-Producing *Escherichia coli* Bacteremia. *Yonsei Med J.* 2014;55(2):467.
163. Kim M, Song KH, Kim CJ, Choe PG, Park WB, Bang JH, et al. Clinical Prediction Score for Community-Onset Bloodstream Infections Caused by Extended-Spectrum Beta-Lactamase-Producing *Escherichia coli* and *Klebsiella* Species. *J Korean Med Sci.* 2019;34(14):e116.
164. Lee CH, Chu FY, Hsieh CC, Hong MY, Chi CH, Ko WC, et al. A simple scoring algorithm predicting extended-spectrum β -lactamase producers in adults with community-onset monomicrobial Enterobacteriaceae bacteremia: Matters of frequent emergency department users. *Medicine (Baltimore).* avr 2017;96(16):e6648.
165. Zahar JR, Lesprit P, Ruckly S, Eden A, Hikombo H, Bernard L, et al. Predominance of healthcare-associated cases among episodes of community-onset bacteraemia due to extended-spectrum β -lactamase-producing Enterobacteriaceae. *Int J Antimicrob Agents.* janv 2017;49(1):67-73.
166. Lin WP, Huang YS, Wang JT, Chen YC, Chang SC. Prevalence of and risk factor for community-

- onset third-generation cephalosporin-resistant *Escherichia coli* bacteremia at a medical center in Taiwan. *BMC Infect Dis.* 12 mars 2019;19(1):245.
167. Lee S, Han SW, Kim KW, Song DY, Kwon KT. Third-generation cephalosporin resistance of community-onset *Escherichia coli* and *Klebsiella pneumoniae* bacteremia in a secondary hospital. *Korean J Intern Med.* 2014;29(1):49.
168. Sung YK, Lee JK, Lee KH, Lee KT, Kang CI. The Clinical Epidemiology and Outcomes of Bacteremic Biliary Tract Infections Caused by Antimicrobial-Resistant Pathogens: *Am J Gastroenterol.* mars 2012;107(3):473-83.
169. Gottesman BS, Shitrit P, Katzir M, Chowers M. Antibiotic Exposure in the Community and Resistance Patterns of *Escherichia coli* Community-Acquired Bloodstream Infection. *Isr Med Assoc J IMAJ.* juin 2018;20(6):382-4.
170. Pedersen G, Schønheyder HC, Steffensen FH, Sørensen HT. Risk of resistance related to antibiotic use before admission in patients with community-acquired bacteraemia. *J Antimicrob Chemother.* janv 1999;43(1):119-26.
171. Ortega M, Soriano A, Marco F, Almela M, Martínez JA, Morata L, et al. Risk factors for the isolation of a third generation cephalosporin resistant strain in patients with community-acquired Enterobacteriaceae bacteraemia. *J Infect.* févr 2016;72(2):268-71.
172. Wi YM, Rhee JY, Kang CI, Chung DR, Song JH, Peck KR. Clinical predictors of methicillin-resistance and their impact on mortality associated with *Staphylococcus aureus* bacteraemia. *Epidemiol Infect.* juill 2018;146(10):1326-36.
173. Lim CJ, Cheng AC, Kong DCM, Peleg AY. Community-onset bloodstream infection with multidrug-resistant organisms: a matched case-control study. *BMC Infect Dis.* 4 mars 2014;14:126.
174. Fan NC, Chen HH, Chen CL, Ou LS, Lin TY, Tsai MH, et al. Rise of community-onset urinary tract infection caused by extended-spectrum β -lactamase-producing *Escherichia coli* in children. *J Microbiol Immunol Infect.* oct 2014;47(5):399-405.
175. Pérez Heras I, Sanchez-Gomez JC, Beneyto-Martin P, Ruano-de-Pablo L, Losada-Pinedo B. Community-onset extended-spectrum β -lactamase producing *Escherichia coli* in urinary tract infections in children from 2015 to 2016: Prevalence, risk factors, and resistances. *Medicine (Baltimore).* déc 2017;96(50):e8571.
176. Søgaard M, Heide-Jørgensen U, Vandenbroucke JP, Schønheyder HC, Vandenbroucke-Grauls CMJE. Risk factors for extended-spectrum β -lactamase-producing *Escherichia coli* urinary tract infection in the community in Denmark: a case-control study. *Clin Microbiol Infect.* déc 2017;23(12):952-60.
177. Tüzün T, Sayın Kutlu S, Kutlu M, Kaleli İ. Risk factors for community-onset urinary tract infections caused by extended-spectrum β -lactamase-producing *Escherichia coli*. *Turk J Med Sci.* 8 août 2019;49(4):1206-11.
178. Boix-Palop L, Xercavins M, Badía C, Obradors M, Riera M, Freixas N, et al. Emerging extended-spectrum β -lactamase-producing *Klebsiella pneumoniae* causing community-onset urinary tract infections: a case-control-control study. *Int J Antimicrob Agents.* août 2017;50(2):197-202.
179. Kung CH, Ku WW, Lee CH, Fung CP, Kuo SC, Chen TL, et al. Epidemiology and risk factors of community-onset urinary tract infection caused by extended-spectrum β -lactamase-producing Enterobacteriaceae in a medical center in Taiwan: A prospective cohort study. *J Microbiol Immunol Infect.* avr 2015;48(2):168-74.
180. Calbo E, Romaní V, Xercavins M, Gómez L, Vidal CG, Quintana S, et al. Risk factors for community-onset urinary tract infections due to *Escherichia coli* harbouring extended-spectrum β -lactamases. *J Antimicrob Chemother.* 1 avr 2006;57(4):780-3.
181. Zhu FH, Rodado MP, Asmar BI, Salimnia H, Thomas R, Abdel-Haq N. Risk factors for community acquired urinary tract infections caused by extended spectrum β -lactamase (ESBL) producing *Escherichia coli* in children: a case control study. *Infect Dis.* 2 déc 2019;51(11-12):802-9.
182. Topaloglu R, Er I, Dogan BG, Bilginer Y, Ozaltin F, Besbas N, et al. Risk factors in community-acquired urinary tract infections caused by ESBL-producing bacteria in children. *Pediatr Nephrol.* mai 2010;25(5):919-25.
183. Yilmaz E, Akalin H, Özbey S, Kordan Y, Sinirtaş M, Gürcüoğlu E, et al. Risk Factors in Community-

- Acquired/Onset Urinary Tract Infections Due to Extended-Spectrum Beta-Lactamase-Producing *Escherichia coli* and *Klebsiella pneumoniae*. *J Chemother.* oct 2008;20(5):581-5.
184. Colodner R, Rock W, Chazan B, Keller N, Guy N, Sakran W, et al. Risk Factors for the Development of Extended-Spectrum Beta-Lactamase-Producing Bacteria in Nonhospitalized Patients. *Eur J Clin Microbiol Infect Dis.* 1 mars 2004;23(3):163-7.
185. Albaramki JH, Abdelghani T, Dalaeen A, Khdaïr Ahmad F, Alassaf A, Odeh R, et al. Urinary tract infection caused by extended-spectrum β -lactamase-producing bacteria: Risk factors and antibiotic resistance. *Pediatr Int.* nov 2019;61(11):1127-32.
186. Guzmán M, Salazar E, Cordero V, Castro A, Villanueva A, Rodulfo H, et al. Multidrug resistance and risk factors associated with community-acquired urinary tract infections caused by *Escherichia coli* in Venezuela. *Biomédica.* 1 mai 2019;39:96-107.
187. Koksall E, Tulek N, Sonmezer MC, Temocin F, Bulut C, Hatipoglu C, et al. Investigation of risk factors for community-acquired urinary tract infections caused by extended-spectrum beta-lactamase *Escherichia coli* and *Klebsiella* species. *Investig Clin Urol.* 2019;60(1):46.
188. Azap ÖK, Arslan H, Şerefhanoglu K, Çolakoğlu Ş, Erdoğan H, Timurkaynak F, et al. Risk factors for extended-spectrum β -lactamase positivity in uropathogenic *Escherichia coli* isolated from community-acquired urinary tract infections. *Clin Microbiol Infect.* févr 2010;16(2):147-51.
189. Rodríguez-Baño J, Alcalá JC, Cisneros JM, Grill F, Oliver A, Horcajada JP, et al. Community Infections Caused by Extended-Spectrum β -Lactamase-Producing *Escherichia coli*. *Arch Intern Med.* 22 sept 2008;168(17):1897.
190. Castillo-Tokumori F, Irey-Salgado C, Málaga G. Worrying high frequency of extended-spectrum beta-lactamase-producing *Escherichia coli* in community-acquired urinary tract infections: a case-control study. *Int J Infect Dis.* févr 2017;55:16-9.
191. Megged O. Extended-spectrum β -lactamase-producing bacteria causing community-acquired urinary tract infections in children. *Pediatr Nephrol.* sept 2014;29(9):1583-7.
192. Chervet D, Lortholary O, Zahar JR, Dufougeray A, Pilmis B, Partouche H. Antimicrobial resistance in community-acquired urinary tract infections in Paris in 2015. *Médecine Mal Infect.* mai 2018;48(3):188-92.
193. Jacmel L, Timsit S, Ferroni A, Auregan C, Angoulvant F, Chéron G. Extended-spectrum β -lactamase-producing bacteria caused less than 5% of urinary tract infections in a paediatric emergency centre. *Acta Paediatr.* janv 2017;106(1):142-7.
194. Brosh-Nissimov T, Navon-Venezia S, Keller N, Amit S. Risk analysis of antimicrobial resistance in outpatient urinary tract infections of young healthy adults. *J Antimicrob Chemother.* 1 févr 2019;74(2):499-502.
195. Colodner R, Kometiani I, Chazan B, Raz R. Risk Factors for Community-Acquired Urinary Tract Infection Due to Quinolone-Resistant *E. coli*. *Infection.* févr 2008;36(1):41-5.
196. Seija V, Fratchez V, Ventura V, Pintos M, González M. Risk factors for community-acquired urinary tract infection caused by fluoroquinolone resistant *E. coli*. *Rev Chil Infectologia Organo Of Soc Chil Infectologia.* août 2014;31(4):400-5.
197. Killgore KM, March KL, Guglielmo BJ. Risk Factors for Community-Acquired Ciprofloxacin-Resistant *Escherichia Coli* Urinary Tract Infection. *Ann Pharmacother.* juill 2004;38(7-8):1148-52.
198. Chaniotaki S, Giakouppi P, Tzouvelekis LS, Panagiotakos D, Kozanitou M, Petrikkos G, et al. Quinolone resistance among *Escherichia coli* strains from community-acquired urinary tract infections in Greece. *Clin Microbiol Infect.* janv 2004;10(1):75-8.
199. Dromigny JA, Nabeth P, Juergens-Behr A, Perrier-Gros-Claude JD. Risk factors for antibiotic-resistant *Escherichia coli* isolated from community-acquired urinary tract infections in Dakar, Senegal. *J Antimicrob Chemother.* 1 juill 2005;56(1):236-9.
200. Hillier S, Roberts Z, Dunstan F, Butler C, Howard A, Palmer S. Prior antibiotics and risk of antibiotic-resistant community-acquired urinary tract infection: a case-control study. *J Antimicrob Chemother.* 1 juill 2007;60(1):92-9.
201. Anesi JA, Lautenbach E, Nachamkin I, Garrigan C, Bilker WB, Wheeler M, et al. Clinical and Molecular Characterization of Community-Onset Urinary Tract Infections Due to Extended-Spectrum

- Cephalosporin-Resistant Enterobacteriaceae. *Infect Control Hosp Epidemiol.* déc 2016;37(12):1433-9.
202. Lee CH, Lee YT, Kung CH, Ku WW, Kuo SC, Chen TL, et al. Risk factors of community-onset urinary tract infections caused by plasmid-mediated AmpC β -lactamase-producing Enterobacteriaceae. *J Microbiol Immunol Infect.* juin 2015;48(3):269-75.
203. Smithson A, Chico C, Ramos J, Netto C, Sanchez M, Ruiz J, et al. Prevalence and risk factors for quinolone resistance among Escherichia coli strains isolated from males with community febrile urinary tract infection. *Eur J Clin Microbiol Infect Dis.* avr 2012;31(4):423-30.
204. Bosch-Nicolau P, Falcó V, Viñado B, Andreu A, Len O, Almirante B, et al. A Cohort Study of Risk Factors That Influence Empirical Treatment of Patients with Acute Pyelonephritis. *Antimicrob Agents Chemother.* déc 2017;61(12):e01317-17, e01317-17.
205. Filiatrault L, McKay RM, Patrick DM, Roscoe DL, Quan G, Brubacher J, et al. Antibiotic resistance in isolates recovered from women with community-acquired urinary tract infections presenting to a tertiary care emergency department. *CJEM.* sept 2012;14(5):295-305.
206. Lim CJ, Cheng AC, Kong DC, Peleg AY. Community-onset bloodstream infection with multidrug-resistant organisms: a matched case-control study. 2014;9.
207. Friedman ND. Health Care-Associated Bloodstream Infections in Adults: A Reason To Change the Accepted Definition of Community-Acquired Infections. *Ann Intern Med.* 19 nov 2002;137(10):791.
208. Kaye KS, Harris AD, Samore M, Carmeli Y. The Case-Case-Control Study Design: Addressing the Limitations of Risk Factor Studies for Antimicrobial Resistance. *Infect Control Hosp Epidemiol.* avr 2005;26(4):346-51.
209. Rodríguez-Baño J, Picón E, Gijón P, Hernández JR, Ruíz M, Peña C, et al. Community-Onset Bacteremia Due to Extended-Spectrum β -Lactamase-Producing Escherichia coli: Risk Factors and Prognosis. *Clin Infect Dis.* janv 2010;50(1):40-8.
210. Zahar JR, Lesprit P, Ruckly S, Eden A, Hikombo H, Bernard L, et al. Predominance of healthcare-associated cases among episodes of community-onset bacteraemia due to extended-spectrum β -lactamase-producing Enterobacteriaceae. *Int J Antimicrob Agents.* janv 2017;49(1):67-73.
211. Hsieh CJ, Shen YH, Hwang KP. Clinical Implications, Risk Factors and Mortality Following Community-onset Bacteremia Caused by Extended-spectrum β -lactamase (ESBL) and non-ESBL Producing Escherichia coli. *J Microbiol Immunol Infect.* juin 2010;43(3):240-8.
212. Søggaard M, Heide-Jørgensen U, Vandenbroucke JP, Schønheyder HC. Risk factors for extended-spectrum β -lactamase-producing Escherichia coli urinary tract infection in the community in Denmark: a casecontrol study. *Clin Microbiol Infect.* 2017;9.
213. Ruppé E, Andremont A. Causes, consequences, and perspectives in the variations of intestinal density of colonization of multidrug-resistant enterobacteria. *Front Microbiol* [Internet]. 2013 [cité 19 oct 2020];4. Disponible sur: <http://journal.frontiersin.org/article/10.3389/fmicb.2013.00129/abstract>
214. Cardoso T, Almeida M, Friedman ND, Aragão I, Costa-Pereira A, Sarmento AE, et al. Classification of healthcare-associated infection: a systematic review 10 years after the first proposal. *BMC Med.* 6 mars 2014;12:40.
215. European Centre for Disease Prevention and Control. Point prevalence survey of healthcare-associated infections and antimicrobial use in European acute care hospitals 2011-2012 [Internet]. 2013. Disponible sur: <https://www.ecdc.europa.eu/en/publications-data/point-prevalence-survey-healthcare-associated-infections-and-antimicrobial-use-0>
216. Ewig S, Welte T, Chastre J, Torres A. Rethinking the concepts of community-acquired and healthcare-associated pneumonia. *Lancet Infect Dis.* avr 2010;10(4):279-87.
217. Kollef MH, Shorr A, Tabak YP, Gupta V, Liu LZ, Johannes RS. Epidemiology and outcomes of health-care-associated pneumonia: results from a large US database of culture-positive pneumonia. *Chest.* déc 2005;128(6):3854-62.
218. Kalil AC, Metersky ML, Klompas M, Muscedere J, Sweeney DA, Palmer LB, et al. Management of Adults With Hospital-acquired and Ventilator-associated Pneumonia: 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clin Infect Dis.* 1 sept 2016;63(5):e61-111.
219. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, et al. Diagnosis and Treatment

- of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med.* 1 oct 2019;200(7):e45-67.
220. Angebault C, Andremont A. Antimicrobial agent exposure and the emergence and spread of resistant microorganisms: issues associated with study design. *Eur J Clin Microbiol Infect Dis.* mai 2013;32(5):581-95.
221. Whitmer GR, Moorthy G, Arshad M. The pandemic *Escherichia coli* sequence type 131 strain is acquired even in the absence of antibiotic exposure. *Leong JM, éditeur. PLOS Pathog.* 19 déc 2019;15(12):e1008162.
222. Morales Barroso I, López-Cerero L, Navarro MD, Gutiérrez-Gutiérrez B, Pascual A, Rodríguez-Baño J. Intestinal colonization due to *Escherichia coli* ST131: risk factors and prevalence. *Antimicrob Resist Infect Control.* déc 2018;7(1):135.
223. Prevel R, Boyer A, M'Zali F, Lasheras A, Zahar JR, Rogues AM, et al. Is systematic fecal carriage screening of extended-spectrum beta-lactamase-producing Enterobacteriaceae still useful in intensive care unit: a systematic review. *Crit Care.* déc 2019;23(1):170.
224. Goulenok T, Ferroni A, Bille E, Lécuyer H, Join-Lambert O, Descamps P, et al. Risk factors for developing ESBL *E. coli*: can clinicians predict infection in patients with prior colonization? *J Hosp Infect.* août 2013;84(4):294-9.
225. Razazi K, Rosman J, Phan AD, Carreaux G, Decousser JW, Woerther PL, et al. Quantifying risk of disease due to extended-spectrum β -lactamase producing Enterobacteriaceae in patients who are colonized at ICU admission. *J Infect.* mai 2020;80(5):504-10.
226. Rottier WC, Bamberg YRP, Dorigo-Zetsma JW, van der Linden PD, Ammerlaan HSM, Bonten MJM. Predictive Value of Prior Colonization and Antibiotic Use for Third-Generation Cephalosporin-Resistant Enterobacteriaceae Bacteremia in Patients With Sepsis. *Clin Infect Dis.* 1 juin 2015;60(11):1622-30.
227. Vodovar D, Marcadé G, Rousseau H, Raskine L, Vicaut E, Deye N, et al. Predictive factors for extended-spectrum beta-lactamase producing Enterobacteriaceae causing infection among intensive care unit patients with prior colonization. *Infection.* août 2014;42(4):743-8.
228. Cho SY, Kang CI, Cha MK, Wi YM, Ha YE, Chung DR, et al. Clinical Features and Treatment Outcomes of Bloodstream Infections Caused by Extended-Spectrum β -Lactamase-Producing *Escherichia coli* Sequence Type 131. *Microb Drug Resist.* août 2015;21(4):463-9.
229. Kim H, Kim YA, Park YS, Choi MH, Lee GI, Lee K. Risk Factors and Molecular Features of Sequence Type (ST) 131 Extended-spectrum β -Lactamase-producing *Escherichia coli* in Community-onset Bacteremia. *Sci Rep.* déc 2017;7(1):14640.
230. Bardini R, Di Carlo S, Politano G, Benso A. Modeling antibiotic resistance in the microbiota using multi-level Petri Nets. *BMC Syst Biol.* nov 2018;12(S6):108.
231. Knight GM, Davies NG, Colijn C, Coll F, Donker T, Gifford DR, et al. Mathematical modelling for antibiotic resistance control policy: do we know enough? *BMC Infect Dis.* déc 2019;19(1):1011.
232. Ruppé E. Causes, consequences, and perspectives in the variations of intestinal density of colonization of multidrug-resistant enterobacteria. *Front Microbiol.* :10.
233. 50M d'euros pour la constitution d'entrepôts de données de santé hospitaliers. Communiqué de presse accessible sur le lien: https://www.health-data-hub.fr/sites/default/files/2022-03/France2030_entrep%C3%B4t%20de%20donn%C3%A9es_VF.pdf.
234. Klann JG, Joss MAH, Embree K, Murphy SN. Data model harmonization for the All Of Us Research Program: Transforming i2b2 data into the OMOP common data model. *PloS One.* 2019;14(2):e0212463.
235. OMOP - Common Data Model. Documentation du SNDS. Accessible sur le lien: <https://documentation-snds.health-data-hub.fr/glossaire/omop.html#origine-verrou-dans-l-utilisation-des-donnees-de-sante-a-l-echelle-internationale>.
236. Découvrir la plateforme Big Data de l'EDS de l'AP-HP. Accessible sur le lien: <https://eds.aphp.fr/nos-services/plateforme-outils>. Consulté le 6 juillet 2022.
237. Santé. CNIL. Accessible sur le lien: <https://www.cnil.fr/fr/sante>. Consulté le 6 juillet 2022.
238. Référentiel relatif aux traitements de données à caractère personnel mis en oeuvre à des fins de

- création d'entrepôts de données dans le domaine de la santé. CNIL. Accessible sur le lien: https://www.cnil.fr/sites/default/files/atoms/files/referentiel_entrepot.pdf. Consulté le 6 juillet 2022.
239. Délibération n° 2021-118 du 7 octobre 2021 portant adoption d'un référentiel relatif aux traitements de données à caractère personnel mis en œuvre à des fins de création d'entrepôts de données dans le domaine de la santé. NOR: CNIL2131763X. JORF n°0249 du 24 octobre 2021. Texte n° 44.
240. Fihn SD, Francis J, Clancy C, Nielson C, Nelson K, Rumsfeld J, et al. Insights From Advanced Analytics At The Veterans Health Administration. *Health Aff (Millwood)*. juill 2014;33(7):1203-11.
241. Price LE, Shea K, Gephart S. The Veterans Affairs's Corporate Data Warehouse: Uses and Implications for Nursing Research and Practice. *Nurs Adm Q*. déc 2015;39(4):311-8.
242. Gradel KO, Nielsen SL, Pedersen C, Knudsen JD, Østergaard C, Arpi M, et al. Seasonal Variation of *Escherichia coli*, *Staphylococcus aureus*, and *Streptococcus pneumoniae* Bacteremia According to Acquisition and Patient Characteristics: A Population-Based Study. *Infect Control Hosp Epidemiol*. août 2016;37(8):946-53.
243. Jansåker F, Holm MKA, Gradel KO, Knudsen JD, Boel JB, Danish Collaborative Bacteraemia Network (DACOBAN), et al. All-cause Mortality Due to Bacteremia during a 60-Day Non-Physician Healthcare Worker Strike. *Clin Infect Dis*. 5 oct 2021;73(7):e1758-61.
244. Pinholt M, Østergaard C, Arpi M, Bruun NE, Schønheyder HC, Gradel KO, et al. Incidence, clinical characteristics and 30-day mortality of enterococcal bacteraemia in Denmark 2006–2009: a population-based cohort study. *Clin Microbiol Infect*. févr 2014;20(2):145-51.
245. Gradel KO, Nielsen SL, Pedersen C, Knudsen JD, Østergaard C, Arpi M, et al. Low Completeness of Bacteraemia Registration in the Danish National Patient Registry. Fernandez-Reyes D, éditeur. *PLOS ONE*. 29 juin 2015;10(6):e0131682.
246. Koch K, Sogaard M, Norgaard M, Thomsen RW, Schonheyder HC, for the Danish Collaborative Bacteremia Network. Socioeconomic Inequalities in Risk of Hospitalization for Community-Acquired Bacteremia: A Danish Population-Based Case-Control Study. *Am J Epidemiol*. 1 mai 2014;179(9):1096-106.
247. Koch K, Nørgaard M, Schønheyder HC, Thomsen RW, Søgaard M, the Danish Collaborative Bacteremia Network (DACOBAN). Effect of Socioeconomic Status on Mortality after Bacteremia in Working-Age Patients. A Danish Population-Based Cohort Study. Trotter CL, éditeur. *PLoS ONE*. 25 juill 2013;8(7):e70082.
248. Rottier WC, Pinholt M, van der Bij AK, Arpi M, Blank SN, Nabuurs-Franssen MH, et al. Attributable mortality of vancomycin resistance in ampicillin-resistant *Enterococcus faecium* bacteremia in Denmark and the Netherlands: A matched cohort study. *Infect Control Hosp Epidemiol*. juin 2022;43(6):719-27.
249. Registre national de la Fédération Hospitalière de France. Accessible sur le lien: <https://etablissements.fhf.fr/>. Consulté le 16/03/22.
250. Lin WP. Prevalence of and risk factor for community-onset third-generation cephalosporin-resistant *Escherichia coli* bacteremia at a medical center in Taiwan. 2019;11.

12 ANNEXES

Annexe 1. Liste des hôpitaux de l'AP-HP

Hôpital	Activité	Nombre de lits* [249]
Albert-Chenevier	SLD & SSR ³¹	463
Ambroise-Paré	MCO ³²	497
Antoine-Béclère	MCO	411
Armand-Trousseau	MCO - pédiatrique	~70 (adultes)
Avicenne	MCO	543
Beaujon	MCO	394
Bicêtre	Mixte	1027
Bichat-Claude Bernard	Mixte	916
Bretonneau	SLD & SSR	202
Broca - La collégiale	SLD & SSR	563
Charles-Foix	SLD & SSR	472
Cochin	MCO	1074
Corentin-Celton	SLD & SSR	505
Dupuytren	SLD & SSR	414
Emile-Roux	SLD & SSR	911
Georges-Clemenceau	SLD & SSR	412
Georges Pompidou	MCO	827
Henri-Mondor	MCO	843
Hôtel-Dieu	MCO	44
Jean-Verdier	MCO	300
La Roche-Guyon	SLD & SSR- pédiatrique	0 (adultes)
Lariboisière et Fernand-Widal	Mixte	976
Louis-Mourier	Mixte	494
Hôpital maritime de Hendaye	SLD & SSR	288
Hôpital maritime de Berck	SLD & SSR	340
Necker-Enfants Malades	MCO - pédiatrique	~200 (adultes)
Paul-Brousse	Mixte	771
Paul-Doumer	SLD & SSR	223
Pitié-Salpêtrière	Mixte	1717
Raymond-Poincaré	Mixte	362
René-Muret	SLD & SSR	492
Robert-Debré	MCO - pédiatrique	~60 (adultes)
Rothschild	SLD & SSR	306
Saint-Antoine	MCO	679
Saint-Louis	MCO	700
Sainte-Périne	SLD & SSR	604
San Salvador	SLD & SSR	351
Tenon	MCO	559
Vaugirard	SLD & SSR	315

Abréviations : MCO : médecine, chirurgie et obstétrique ; SLD : soins de longue durée ; SSR : soins de suite et réadaptation. *Lits adultes

³¹ Séjours de longue durée et Soins de suite et de réadaptation

³² Soins aigus : médecine, chirurgie, obstétrique

Annexe 2. Liste des bactéries et champignons inclus dans la base BactHub

Cocci Gram positifs :

- . *Staphylococcus aureus* ou *lugdunensis* ou *Staphylococcus* coagulase négative.
- . *Streptococcus* beta hémolytique (A B (dont *agalactiae*) C G), D (*gallolyticus* et *bovis*), pneumocoque, autres streptocoques.
- . *Enterococcus faecium* ou *faecalis*.
- . Cocci Gram positifs anaérobies : *Peptostreptococcus*.

Cocci Gram négatifs : *Branhamella* (ou *Moraxella*) *catarrhalis*, *Neisseria meningitidis* ou *gonorrhoeae*.

Bacilles Gram positifs :

- . Aérobie : *Listeria monocytogenes*, *Bacillus anthracis*, *Nocardia*.
- . Anaérobies : *Clostridium perfringens*, *Propionibacterium acnes*, *Actinomyces*.

Bacilles Gram négatifs :

- . Entérobactéries : *Salmonella spp*, *Escherichia coli*, *Proteus mirabilis* et non mirabilis, *Klebsiella pneumoniae* et *oxytoca*, *Citrobacter spp*, *Klebsiella aerogenes*, *Enterobacter cloacae* ou *aerogenes*, *Serratia spp*, *Providencia spp*, *Morganella spp*, *Citrobacter freundii*, *Hafnia alvei*,
- . Petits Bacilles Gram négatifs : *Pasteurella multocida*, *Haemophilus spp*, *Bordetella spp*, *Legionella spp*, *Vibrio spp*, *Brucella spp*, *Rhodococcus equi*, bactéries du groupe HACEK (*Haemophilus spp*, *Actinobacillus spp*, *Cardiobacterium spp*, *Eikenella spp*, *Kingella spp*), *Campylobacter spp*, *Francisella tularensis*.
- . Bacilles Gram négatifs anaérobies : *Bacteroides fragilis* et *spp*, *Fusobacterium spp*.
- . Autres : *Acinetobacter baumannii*, *Pseudomonas spp*, *Stenotrophomonas maltophilia*.

Champignons : *Candida* et *Fusarium*.

Exclus : bactéries intracellulaires (*Chlamydia*, *Mycoplasme*, *Rickettsia*, *Coxiella*, *Mycobacterium leprae*) ou bactéries spirochètes (*Treponema*, *Borrelia*, *Leptospira*).

Annexe 3. Étapes de nettoyage et structuration de la base

Séjours

Sélection des séjours exploitables et dédoublement :

- Sélection des séjours exploitables (date de début antérieure à la date de fin, date de fin \leq 31/12/2020) ou potentiellement exploitables (date de début \leq 31/12/2020 et date de fin inexistante).
- Suppression des séjours avec des dates non plausibles (date de début à 1800, etc.).
- Suppression des séjours non liés à des données exploitables (seuls sont gardés les séjours ayant au moins une donnée disponible de type codes CIM-10, actes de la CCAM, observations physiologiques, résultats de laboratoire, résultats de microbiologie, documents médicaux, données médicamenteuses).

- **Algorithme détectant les doublons imbriqués, ou qui se chevauchent.**
 - Pour un même patient, deux séjours (avec sd = date de début, ed = date de fin) se chevauchent si et seulement si (1) $sd1 \leq sd2$ et (2) $sd2 < ed1$ et (3) $sd1, ed1, sd2, ed2$, sont non nuls.
 - Pour un même patient, deux séjours sont imbriqués si et seulement si (1) $sd2$ compris entre $sd1$ et $ed1$, et (2) $ed2 \leq ed1$ ou $ed2$ est à null.

- **Algorithme de dédoublement** permettant de dédoubler les séjours et leurs données dans toutes les tables de la base.

- Les règles de dédoublement sont différentes en fonction du type d'hospitalisation (I = hospitalisation complète, II = hospitalisation incomplète, III = pré-hospitalisation, O = consultation externe, U = urgences) et de la source des données (orbis = logiciel ORBIS, hegp = données de l'HEGP, glims = logiciel GLIMS) des séjours et des doublons. Pour chaque cas, ces règles ont été déterminées après avoir évalué la qualité et la présence de données exploitables.

- Règles pour les séjours / doublons imbriqués :
 - orbis / glims** = remplacer le numéro de séjour du doublon par celui du séjour dans la table des données de microbiologie, supprimer le doublon de la table avec les données de séjour.
 - glims / orbis** = supprimer le séjour et le doublon.
 - orbis / orbis** de type **I / O, I / U** = remplacer le numéro de séjour du doublon par celui du séjour dans toutes les tables, supprimer le doublon de la table des données de séjour.
 - orbis / orbis** de type **II / I** = supprimer le séjour.
 - orbis / orbis** de type **II / O, O / II, U / I, II / II, II / U, I / I, O / I, I / II** = supprimer le séjour et le doublon.
 - glims / glims** = supprimer le séjour et le doublon.
 - hegp / hegp, orbis / hegp, hegp / orbis, glims / hegp, hegp / glims** = supprimer le séjour et le doublon.

Règles pour les séjours / doublons se chevauchant :

- orbis / glims** : remplacer le numéro de séjour du doublon par celui du séjour dans la table des données de microbiologie, supprimer le doublon de la table des données de séjour.
- glims / orbis** : supprimer le séjour et le doublon.
- orbis / orbis** de type **I / II, I / O, I / U** : supprimer le doublon.
- orbis / orbis** de type **II / II, I / I** : supprimer le séjour et le doublon.
- orbis / orbis** de type **II / I, O / I, U / I** : supprimer le séjour.
- glims / glims** : supprimer le séjour et le doublon.
- hegp / hegp, hegp / orbis, glims / hegp, orbis / hegp** : supprimer le séjour et le doublon.

Définition de trois fonctions pour le dédoublonnage :

Fonction delete_from_visit : suppression des séjours selon les règles pré-citées.

Fonction replace_in_tables : pour la table appelée par la fonction, remplace le numéro de séjour du doublon par le numéro de séjour du séjour principal.

Fonction dedoublonnage : appelle la fonction replace_in_tables pour une table et un type de doublons particulier.

Sélection des séjours d'inclusion avec bactériémie :

- Sélection des séjours (1) de patients majeurs, (2) avec une date de début \leq 31/12/2019, et une date de fin entre le 1/1/2016 et le 31/12/2019, (3) dont la source est le logiciel ORBIS ou les données de l'HEGP, et (4) avec des données de bactéries dans le sang, et / ou un code cim10 de bactériémie (A40, A41, A32.7, A39.4, A42.7, A02.1, A54.8, A48.0) en DP, DR ou DAS, en ne prenant que les codes du PMSI AREM.

Nettoyage des données des séjours :

- Sélection des séjours de patients avec au moins un séjour d'inclusion avec bactériémie.
- Ajout d'une colonne indiquant si le séjour est ou non un séjour d'inclusion avec bactériémie.
- Suppression des séjours avec exclusivement des données GLIMS.
- Suppression des consultations.

Patients

Nettoyage de la table des données patients :

- Sélection des patients avec un moins un séjour d'inclusion avec bactériémie.
- Correction de la variable décès.

Données de microbiologie

Nettoyage et structuration des données de microbiologie :

- Catégorisation des milieux de prélèvement (sang, urines, etc.)
- Fusion des données liées à un même prélèvement, qui étaient réparties sur plusieurs lignes différentes dans la table.
- Sélection des codes de bactéries et champignons utiles à l'étude (MBIO:1.*, MBIO:2CND.*, ^MBIO:2_LEVCAL\$, ^MBIO:2_LEVNCAL\$, MBIO:STAPHY ,MBIO:ENTCS , MBIO:CANDIDA, MBIO:STPAUT, MBIO:BGNANA, MBIO:CGPANA, MBIO:STCGNG, MBIO:STREPT) et suppression des prélèvements identifiés à tort comme des bactéries (MBIO:1BIOMOL, MBIO:1_INCONNU). Les codes de type MBIO:1.* correspondent à des bactéries, sauf exception ; les codes de type MBIO:2.* correspondent à des levures ; les codes de type MBIO: sans chiffre correspondent à des micro-organismes non catégorifiés, sélectionnés car utiles à l'étude.
- Transformation des données de l'antibiogramme en colonnes : elles étaient regroupées dans une seule colonne sous la forme d'un xml. Le xml est parsé pour pouvoir extraire les colonnes suivantes : nombre de colonies, libellé du matériel, libellé du germe, code du matériel, antibiogramme, nombre d'antibiotiques testés, présence d'une BMR, présence d'une BHR.
- Réassignation des observations de microbiologie aux séjours dont les dates leurs sont conformes, pré-dédoublonnage.
 - Sélection des prélèvements qui ont été couplés au mauvais séjour (date de prélèvement en dehors des dates de séjour) mais qui sont assignables à un autre séjour avec une correspondance parfaite des dates.
 - Si l'observation est assignable à un séjour unique, on l'assigne.
 - Si l'observation est assignable à plusieurs séjours, on l'assigne en priorité à une hospitalisation complète, puis à un séjour aux urgences, puis à une consultation externe, et enfin à une hospitalisation incomplète qui est le type de séjour le moins fiable. S'il n'y a aucun de ces types de séjours qui est renseigné, on assigne l'observation par défaut au premier séjour de la liste, parmi les séjours assignables.

- On corrige les numéros de séjours des observations selon les assignations qui ont été faites dans la table de microbiologie.
- Organisation de l'antibiogramme en ajoutant des colonnes de résistance pour chaque antibiotique, avec un nom normalisé.

Épisodes de bactériémie

Regroupement des bactériémies en épisodes et structuration des données des autres tables à l'échelle des épisodes :

- **Algorithme de détection des épisodes :**
 - Détection des épisodes de bactériémie basés sur les données d'hémocultures :
Après avoir trié la table des données de microbiologie selon les numéros de patients et la date de l'hémoculture : si deux prélèvements qui se suivent ont les mêmes patients et ont eu lieu à 72h d'intervalle ou moins, on les regroupe dans le même épisode. Sinon, on les classe dans 2 épisodes différents. Si les hémocultures d'un même épisode sont rattachées à plusieurs séjours différents, l'épisode est lié à l'ensemble de ces séjours.
 - Détection des épisodes basés sur les données CIM10 :
En l'absence de donnée microbiologique, nous considérons chaque séjour ayant un diagnostic CIM10 de bactériémie comme étant un épisode à part entière.
- Création des dates de début et de fin des épisodes
 - Pour les épisodes de bactériémie basés sur les données d'hémocultures, la date de début est la date du 1er prélèvement positif, et la date de fin est la date du dernier prélèvement + 72h.
 - Pour les épisodes basés sur les données CIM10, nous avons pris les dates de début et de fin de séjour comme dates de début et de fin de l'épisode puisque nous n'avons pas l'information de quand a été détectée la bactériémie.
- Création des variables de mortalité
 - A 30 jours : 1 si la date de décès du patient est à 30 jours ou moins de la date de début de l'épisode, 0 sinon.
 - A 90 jours : 1 si la date de décès du patient est à 90 jours ou moins de la date de début de l'épisode, 0 sinon.
 - Pendant le séjour : 1 si le patient meurt pendant le séjour hospitalier associé à l'épisode, 0 sinon.
- **Algorithme de classification des épisodes issus des données de microbiologie en communautaires ou nosocomiaux :**
Après jointure de la table des épisodes avec la table contenant tous les séjours de la base (d'inclusion, ou de non-inclusion) :
On parcourt la table après l'avoir triée par numéro de patient et par date de début de séjour.
Dès qu'on rencontre un séjour associé à un épisode on vérifie s'il remplit les critères pour être considéré comme communautaire :
 - Délai < 48h entre le début du séjour et le début de l'épisode de bactériémie.
 - Et > 7 jours entre la date de fin d'un précédent séjour et la date de début du séjour où a lieu l'épisode.
 Quand plusieurs séjours étaient associés à un même épisode, c'est la date de début du premier de ces séjours qui est considérée, en ignorant les séjours pour lesquels la date de début de l'épisode est en dehors des dates de séjour avec un écart supérieur à 24h. Certains épisodes à séjours multiples n'ont pas été classifiés car aucun séjour n'a été jugé suffisamment fiable.
- Jointure de la table des épisodes à toutes les autres tables :
 - Table des données de microbiologie :
 - Liaison sur la base du numéro de séjour.

- Identification par une variable des observations de microbiologie prélevées pendant les dates de l'épisode, et de celles prélevées pendant le séjour lié à l'épisode, mais en dehors des dates de l'épisode.
- **Algorithme identifiant les contaminants dans le sang :**
 - Espèces potentiellement contaminantes : *Corynebacterium sp*, *Staphylococcus epidermidis*, *Staphylococcus pettenkoferi*, *Staphylococcus haemolyticus*, *Staphylococcus hominis*, *Staphylococci à coagulase négative*, *Staphylococcus capitis*, *Staphylococcus hominis subsp. hominis*, *Micrococcus luteus*, *Staphylococcus capitis subsp. capitis*, ou *Staphylococcus warneri*.
 - Au sein d'un même épisode, défini à partir des données de microbiologie, nous avons considéré qu'une hémoculture à l'une de ces espèces n'était pas une contamination si les 3 conditions suivantes étaient réunies : (1) présence d'au moins une autre hémoculture positive à la même espèce dans les 72h ; (2) présence pour les deux hémocultures d'au moins 6 résultats de sensibilité aux antibiotiques sur 8 disponibles pour les antibiotiques suivants : l'oxacilline, la céfoxitine, la tobramycine, la gentamicine, la rifampicine, l'érythromycine, la clindamycine, le cotrimoxazole ; (3) similarité entre les deux hémocultures des résultats des tests de sensibilité disponibles pour ces antibiotiques.
- Table des passages dans des unités d'hospitalisation au cours d'un séjour :
 - Organisation et normalisation.
 - Liaison avec les épisodes sur la base du numéro de séjour.
 - Identification par une variable des passages effectués pendant les dates de l'épisode, et de ceux pendant le séjour lié à l'épisode, mais en dehors des dates de l'épisode.
- Table des résultats de laboratoire, et des observations physiologiques :
 - Liaison avec les épisodes sur la base du numéro de séjour.
 - Pour chaque variable d'intérêt (créatininémie, leucocytes, pression artérielle moyenne, etc.) : on détermine les bornes normales de la variable, et on indique comment détecter les pires valeurs via les variables booléennes MIN et MAX (exemple : la pire valeur de la créatininémie est la valeur la plus élevée).
 - **Algorithme de détection des variables d'intérêt, et leur pire valeur dans une période définie (+/-24h, +/-48h, +/-72h autour du début de l'épisode, ou durant le séjour).**
- Autres tables (documents médicaux, codes CIM-10, actes de la CCAM) : liaison sur la base du numéro de séjour.
- Structuration des épisodes :
 - Exclusion des épisodes avec que des contaminants.
 - Exclusion des épisodes sans hospitalisation complète ou ne passant pas une nuit à l'hôpital (à l'exception de ceux qui sont passés par une unité de réanimation, de médecine, de chirurgie, ou en hôpital de semaine, et sont morts dans les 24h).
 - Reclassement des épisodes issus d'un transfert en nosocomial.
 - Ajout d'une variable calculant la durée depuis le début de l'épisode jusqu'à la fin du séjour (en cas de séjours multiples c'est le dernier séjour qui est considéré pour la date de fin)
 - Création des variables de début et de fin de séjour pour chaque épisode. Dans le cas où un épisode est lié à plusieurs séjours, la date de début est celle du premier séjour, et la date de fin celle du dernier séjour.
 - Création de variables d'intérêt grâce aux codes CIM-10 et aux actes de la CCAM réalisés pendant les séjours reliés aux épisodes (par exemple : cim10_diabete (liste des codes liés au diabète présents pendant le séjour lié à l'épisode), cim10_is_diabete (1/0),

cim10_codes_infections (liste des codes d'infection bactérienne présents pendant le séjour lié à l'épisode, hors bactériémie), cim10_choc_septique (1/0), ccam_cathe_veineux_central (1/0, présence ou non d'un acte de poste de cathéter veineux central pendant le séjour lié à l'épisode), etc.).

- A partir des données des passages dans les unités d'hospitalisation, création du parcours de soin des patients pendant le séjour lié à l'épisode, en regroupant les unités d'hospitalisation dans des catégories : urgences, médecine, chirurgie, hôpital de semaine, réanimation, soins intensifs, etc.
- Création de variables identifiant le passage en réanimation (1/0) ou en soins intensifs (1/0) pendant le séjour lié à l'épisode et la durée de ce passage.
- Création de variables identifiant le passage en réanimation (1/0) ou en soins intensifs (1/0) pendant l'épisode, de manière stricte (donc pendant les dates de positivité des hémocultures), et la durée de ce passage.

Données médicamenteuses

Structuration des données issues du module ORBIS médicament :

- Fusion des données réparties sur plusieurs lignes (il y avait 3 lignes par prescription)
- Filtrage en nous limitant aux codes ATC d'intérêt : J01 (antibactériens à usage systémique), J02 (antimycosiques à usage systémique), H02 (corticoïdes à usage systémique), C01CA (agents adrénergiques et dopaminergiques).
- Pré-nettoyage de la colonne « modalités » stockant les informations de posologies : suppression des espaces en début et fin de phrase, suppression des espaces en trop, suppression des modalités incomplètes ou avec des points d'interrogation (exemple : ? mg le matin), normalisation des modalités (exemple : Bolus initial : 1200 mg || Dose : 2000 mg/j devient Bolus initial : 1200 mg puis 2000 mg/j ; 10 mg/kg (soit 680 mg) devient 680 mg), raccourcissement des modalités trop longues sans informations utiles, suppression des informations non utilisées ou de certains mots, suppression des virgules en dehors des nombres décimaux, remplacement des virgules des nombres décimaux par des points)
- Extraction des données de posologie via la colonne « modalités » : la colonne « modalités » est parsée grâce à des règles d'expression régulières. Chaque modalité possède sa règle. Nous avons recensé 14 modalités couvrant 98% des données.

Tableau récapitulatif des types de posologies et leur découpage :

	type poso	exemple	dose_recue	unite_prise	frequence
	prise_seul	1 g	1	g	pu
	freq	1,7 g/j	1,7	g	j
	fois	12 g (en 6 fois)	72	g	pm
	jours1	2 g lundi mercredi vendredi	2	g	3/7
	jours2	2 MU/j lundi mercredi	2	g	2/7
	bolus	Bolus initial : 1 mg puis 1 mg/j	1	mg	pu
	bolus (suite)	Bolus initial : 1 mg puis 1 mg/j	1	mg	j
	tous jours1	2 cpr tous les 2 jours	2	cpr	1/2
	tous jours2	2 cpr 2 cpr (soit 4 cpr/j) tous les 14 jours	4	cpr	1/14
	toutes heures	1 cpr toutes les 48 heures	1	cpr	1/2
	prise2	1 g par prise	1	g	j
	prise3	10 mg/kg (soit 790 mg par prise)	790	mg	j
	sur1	1 cpr 1 jour sur 2	1	cpr	1/2
	sur2	500 g 500 g (soit 1000 g/j) 1 jour sur 2	1000	g	1/2

pu = prise unique ; pm = prise multiple ; 1/2 = 1 jour sur 2 ; 2/7 = 2 fois par semaine

Une expression régulière est créée pour chaque modalité, permettant l'extraction des données de posologie.

- Correction de la dose reçue et de l'unité de prise (g, mg, etc.) grâce aux informations des noms des médicaments : les données non précises du type 1 cpr (1 comprimé) sont remplacées par leur équivalent en unité de masse conventionnelle (g, mg, etc.) et en récupérant la dose par comprimé. Voici la liste des unités modifiées : sachet, cpr, gel, Flac, poche, amp, Flac_inj, SER_P, C_MES, AMPB, SER.
- Calcul du délai entre la date de prescription et la date de début de l'épisode.
- Normalisation des noms de médicaments et des voies d'administration.

Structuration des données issues du logiciel Phedra :

- Dédoublonnage et nettoyage des données pour ne conserver que les données sur les médicaments d'intérêt. Nous conservons le nombre de répétitions des lignes dans la table brute, car après l'exploration il s'avère que les doublons correspondent à des administrations différentes d'une même prescription.
- Liaison entre les données Phedra et les données de séjour sur le numéro de patient et les dates de séjour, car les numéros de séjours des données Phedra est incorrect et différent des numéros de séjours des autres tables. Pour cela, après jointure, on filtre sur l'un des critères suivants :
 - Date de début de prescription comprise dans les dates du séjour.
 - Date de fin de prescription comprise dans les dates du séjour.
 - Les dates de prescription sont toutes deux comprises dans les dates du séjour.
 - Les dates de prescription englobent les dates du séjour.
- Jointure avec les numéros d'épisodes, sur les numéros de séjours corrigés.
- Enrichissement des données Phedra grâce aux codes CIP13 et UCD afin d'avoir les doses et unités des médicaments.
- Correction des données de médicaments ayant une unité de type : gelule, gel, flacon, cpr, pulv, tube, inh, dose, mesur, cuilm, sachet, bouff, gt, poche, fois, cp, pose, suivi, inj, cuilc, cuils, app. Ces unités sont remplacées par des unité de masse conventionnelles et leurs doses sont recalculées.

Annexe 4. Nomenclatures

Codes CIM-10 utilisés pour sélectionner et classifier les comorbidités, selon Quan et al. [107]	
Infarctus du myocarde	I21, I22, I25.2
Insuffisance cardiaque	I09.9, I11.0, I25.5, I13.0, I13.2, I50, I42.0, I42.5, I42.6, I42.7, I42.8, I42.9, I43.0, P29.0
Maladie vasculaire périphérique	I70, I71, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9
Maladie cérébrovasculaire	G45, G46, H34.0, I60, I61, I62, I63, I64, I65, I66, I67, I68, I69
Hémiplégie / paraplégie	G04.1, G11.4, G80.1, G80.2, G81, G82, G83.0, G83.1, G83.2, G83.3, G83.4, G83.9
Démence	F00, F01, F02, F03, F05.1, G30, G31.1
Maladie pulmonaire chronique	I27.8, I27.9, J40, J41, J42, J43, J44, J45, J46, J47, J60, J61, J62, J63, J64, J65, J66, J67, J68.4, J70.1, J70.3
Maladie systémique	M05, M06, M31.5, M32, M33, M34, M35.1, M35.3, M36.0
Pathologie ulcéreuse	K25, K26, K27, K28
Diabète non compliqué	E10.0, E10.1, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9
Diabète avec complication chronique	E10.2, E10.3, E10.4, E10.5, E10.7, E11.2, E11.3, E11.4, E11.5, E11.7, E12.2, E12.3, E12.4, E12.5, E12.7, E13.2, E13.3, E13.4, E13.5, E13.7, E14.2, E14.3, E14.4, E14.5, E14.7
Maladie rénale	I12.0, I13.1, N03.2, N03.3, N03.4, N03.5, N03.6, N03.7, N05.2, N05.3, N05.4, N05.5, N05.6, N05.7, N18, N19, N25.0, Z49.0, Z49.1, Z49.2, Z94.0, Z99.2
Tumeur maligne	C00, C01, C02, C03, C04, C05, C06, C07, C08, C09, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, C25, C26, C30, C31, C32, C33, C34, C37, C38, C39, C40, C41, C43, C45, C46, C47, C48, C49, C50, C51, C52, C53, C54, C55, C56, C57, C58, C60, C61, C62, C63, C64, C65, C66, C67, C68, C69, C70, C71, C72, C73, C74, C75, C76, C97
Lymphome	C81, C82, C83, C84, C85, C88, C90, C96
Leucémie	C91, C92, C93, C94, C95
Cancer	Codes tumeur maligne + lymphome + leucémie
Tumeur solide métastatique	C77, C78, C79, C80
Maladie hépatique légère	B18, K70.0, K70.1, K70.2, K70.3, K70.9, K71.3, K71.4, K71.5, K71.7, K73, K74, K76.0, K76.2, K76.3, K76.4, K76.8, K76.9, Z94.4
Maladie hépatique modérée / sévère	I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7
SIDA	B20, B21, B22, B24

Codes CIM-10 de portage ou d'infection à bactérie multirésistante ou hautement résistante		
Acronyme	Nom	Codes CIM-10
SARM	<i>Staphylococcus aureus</i> résistant à la méthicilline	U8210, U82100, U8210+0, U82101, U8210+1
BMR	Bactérie multirésistante aux antibiotiques	U8371, U83710, U8371+0, U83711, U8371+1
BHR	Bactérie hautement résistante	U8370, U83700, U8370+0, U83701, U8370+1
BLSE	Bactérie productrice de bêta-lactamases à spectre élargi	U822, U8220, U822+0, U8221, U822+1

Codes CIM-10 utilisés pour classier les infections bactériennes, adaptés d'Opatowski et al.[29]

Infection bactérienne	Codes CIM-10
Bactériémie	A40.0, A40.1, A40.2, A40.3, A40.8, A40.9, A41.0, A41.1, A41.2, A41.3, A41.4, A41.5, A41.8, A41.9, A32.7, A39.4, A42.7, A02.1, A54.8, A48.0
Digestive	A00, A01, A02, A03, A04, A05, A22.2, A39.1, A42.1, D73.3, K29.1, K29.5, K29.6, K29.7, K29.8, K29.9, K35, K36, K37, K40.1, K40.4, K41.1, K41.4, K42.1, K43.1, K43.4, K43.7, K44.1, K45.1, K57.0, K57.2, K57.4, K57.8, K61, K63.0, K65, K67, K75.0, K80.0, K80.1, K80.3, K80.4, K81, K83.0, K83.8
Ostéoarticulaire	M00, M01.0, M01.2, M01.3, M46.2- M46.5, M49.1-M49.3, M86, Z76.800
Cardio-vasculaire	A39.5, I30.1, I30.9, I32.0, I33, I40.0, I41.0, I70.01, I70.21, I70.81, I70.91, I98.0, J98.5
Congénitale ou Périnatale	A50, O23, O41.1, O85, O86, O88.3, O91, O98.1, O98.2, O98.8, O98.9, P00.2, P23.1, P23.2, P23.3, P23.4, P23.5, P23.6, P23.8, P36, P37.2, P37.8, P39.0, P39.2, P39.8, P39.9, P78.1
Sur matériel	T85.7, T84.5, T84.6, T84.7, T82.6, T82.7, T83.5
Oculaire ou oto-rhino-laryngée	A36.0, A36.1, A36.2, A42.2, A74.0, H04.3, H04.4, H10.0, H10.3, H10.4, H10.5, H13.1, H15, H15.0, H15.1, H16, H16.0-H16.9, H30, H30.0-H30.9, H44.0, H60, H62.0, H66, H67, H67.0, H68.0, H70, H71, H73, H75, J01-J06, J32, J34.0, J35.0, J36, J37, J39.0, J39.1, K04.4, K04.5, K04.6, K04.7, K05.0, K05.1, K05.2, K05.3, K10.2, K10.3, K11.3, K12.2
Génitale	N45.0, N45.9, N49, N51, N70, N73, N75, N76, N77
Respiratoire basse	A22.1, A31.0, A37, A42.0, A43.0, A70, J13-J15, J16, J17.0, J17.8, J18, J20, J20.0-J20.2, J20.8, J20.9, J22, J40, J41, J42, J85, J86, J90
Muscle	M60.0
Système nerveux	A32.1, A39.0, G00, G01, G03.9, G04.2, G05.0, G06, G07
Autres	A22.8, A22.9, A23, A24, A25, A26.8, A26.9, A27, A28, A30, A31.8, A31.9, A32.8, A32.9, A33, A34, A35, A36.8, A36.9, A39.8, A39.9, A42.8, A42.9, A43.8, A43.9, A44.0, A44.8, A44.9, A48, A49, A65, A66, A67, A68, A69.1, A69.2, A69.8, A69.9, A74.8, A74.9, A75, A77, A78, A79, L04, R02, T814
Sexuellement transmissible	A51, A52, A53, A54, A55, A56, A57, A58, A71
Peau & tissus mous	A20, A21, A22.0, A26.0, A31.1, A32.0, A36.3, A38, A43.1, A44.1, A46, A69.0, L00-L03.9, L05.0, L08, L30.3, L73.2, L88, L89-L89.9, L97, M65.0, M71.0, M71.1, M72.6, N61, T31, T79.3, T80.2, T87.4, T88.0
Urinaires	N08.0, N10, N11.0, N11.1, N12, N13.6, N15.1, N16.0, N29.0, N29.1, N30.0, N30.2, N30.3, N30.8, N30.9, N34, N34.0-N34.2, N39.0, N41.0-N41.3

Codes CIM-10 de choc septique et d'arrêt cardiaque

Choc septique	R572
Arrêt cardiaque	I469, R092, P298, I978

Codes des actes de la CCAM sélectionnés

Acte	Code CCAM
Cathéter artériel	ENLF001, EBLA001, EDLA001, EDLF002
Cathéter veineux central	EPLB002, EPLF002, EPLF005, EPLF004, EPLF003, EBGA001, EBLA002, DHLF001, EBLA003, EPLA002, EPPP003, EGLF001, EHLA001, EPPP002, EQQP011
Ventilation invasive	GELD004, GELD002, GELD003, GLLD008, GLLD015, GLLD007, GLLD004, GLLD006, GLLD013, GLLD009, GLLD010, GLLD011, DKMD001, DKMD002
Ventilation non invasive	GLLD012, GLLD021, GLLD019, GLLD003, GLLD002, GLLD001
Support par amines vasoactives	EQLF001, EQLF003
Épuration extra rénale	JVJB002, JVJF003, JVJF005, JVJF004, JVJF006, JVJF007, JVJF008, JVJF002, JVJB001

Annexe 5. Comparaison des identifications de bactériémies-bactéries et de sites primaires dans les CRH de 300 épisodes, entre le programme de TAL et une lecture experte, et classification

		Lecture experte	
		Présence	Absence
Identification de bactériémies-bactéries			
Présence	Prédiction complète, sans FP = 186 (VP)		
	Prédiction complète, avec FP = 27 (FP)*		3 (FP)*
	Prédiction partielle, sans FP = 3 (FN)**		
Absence	17 (FN)**		64 (VN)
Identification de sites primaires[#]			
Présence	Prédiction complète, sans FP = 93 (VP)		
	Prédiction complète, avec FP = 8 (FP) [§]		7 (FP) [§]
	Prédiction partielle, sans FP = 11 (FN) [@]		
	Prédiction fautive = 5 (FP) ^{§,@}		
Absence	52 (FN) [@]		Pas de bactériémie = 67 (VN) Bactériémie sans site primaire = 57 (VP)

[#]Référence pour les bactériémies-bactéries = lecture experte.

* Causes de FP :

- Bactérie d'une infection sans bactériémie, présentée dans la même phrase qu'une bactériémie, n = 12 (exemple : « Nouveau sepsis le 31/10/2015 à point de départ probable pulmonaire à *Klebsiella Pneumoniae*, *Staphylococcus aureus* méticilline sensible et des levures et une hémoculture positive à *Pseudomonas Aeruginosa* (translocation digestive sur fécalome ?) »).
- Bactérie d'une infection sans bactériémie, liée à une bactériémie par une bactérie, n = 8 (exemple : « bactériémie à *Escherichia coli* », et « angiocholite à *Escherichia coli* et *Klebsiella pneumoniae* »).
- Erreur de sectionnement du CRH, n = 3.
- Absence de point en fin de phrase, n = 2.
- Interprétation d'expert nécessaire pour l'identification correcte de la(des) bactérie(s), n = 2 (exemple : « hémoculture positive à BGN, qui s'est avérée finalement positive à *Staphylococcus haemolyticus* »).
- Phrase listant des examens biologiques réalisés dont une hémoculture et une antigénurie Légionnelle / « Antigénémie urinaire Légionnelle », n = 2.
- Non-identification de la négation, n = 1 (« *Haemophilus*, avec hémoculture négative »).

**Causes de FN :

- Erreur de sectionnement du CRH, n = 8.
- Variations de morphologie non listées dans la terminologie, n = 6 (exemple : *Actinobacter baumannii* au lieu de *Acinetobacter baumannii* ; *Kl. Pneumoniae* au lieu de *K. pneumoniae* ; Staph à coag negative et *s. haemoliticus* au lieu de *Staphylocoque coagulase négative* et *S. haemolyticus*).
- Saut de ligne, n = 3.
- Indice de négation dans la même phrase, n = 2.
- Interprétation d'expert nécessaire, n = 1.

[§]Causes de FP :

- Interprétation d'expert nécessaire, n = 5 (exemple : suspicion d'endocardite, finalement éliminée).
- Erreur de sectionnement du CRH, n = 4.
- Site infectieux lié à une infection sans bactériémie, citée dans la même phrase qu'une bactériémie, n = 4.
- Présence d'une bactérie commune avec une bactériémie, n = 3.
- Variations de morphologie menant à une fausse identification, n = 2 (exemple : *K. pneumonia* au lieu de *K. pneumoniae* a mené à l'identification de *pneumonia* comme une pneumonie).
- Interprétation d'expert nécessaire, n = 1 (exemple : « si infection pulmonaire » évoque une possibilité non confortée par des éléments objectifs).
- Indice de négation non listé dans la terminologie, n = 1.

[@]Causes de FN :

- Interprétation d'expert nécessaire, n = 37.
- Variations de morphologie non listées dans la terminologie, n = 16 (exemple : « sepsis biliaire »).
- Erreur de sectionnement du CRH, n = 10.
- Non-identification à tort, n = 5 (dont 4 liées aux infections sur cathéter central).

Annexe 6. Comparaison des identifications de sites primaires d'épisodes de bactériémie entre le programme de TAL ou les codages CIM-10, et une lecture experte, et classification

		Lecture experte	
		Présence	Absence
Programme de TAL			
Identification de sites primaires dans les CRH de 300 épisodes*			
Présence	Prédiction complète, sans FP = 90 (VP)	6 (FP)	
	Prédiction complète, avec FP = 7 (FP)		
	Prédiction partielle, sans FP = 2 (FN)		
	Prédiction fausse = 5 (FP)		
Absence	48 (FN)	58 (VP)	
Identification de sites primaires dans les CRH de 216 épisodes***			
Présence	Prédiction complète, sans FP = 90 (VP)	6 (FP)	
	Prédiction complète, avec FP = 7 (FP)		
	Prédiction partielle, sans FP = 2 (FN)		
	Prédiction fausse = 5 (FP)		
Absence	48 (FN)	58 (VP)	
Codages CIM-10			
Identification de sites primaires pour 216 épisodes***			
Présence	Prédiction complète, sans FP = 56 (VP)	36 (FP)	
	Prédiction complète, avec FP = 33 (FP)		
	Prédiction partielle, sans FP = 3 (FN)		
	Prédiction partielle, avec FP = 5 (FP)		
	Prédiction fausse = 28 (FP)		
Absence	29 (FN)	26 (VP)	

Abréviations : CRH : comptes-rendus d'hospitalisation. *Référence pour les bactériémies-bactéries = microbiologie.

***Les 216 épisodes ont des CRH avec une information complète sur les bactériémies-bactéries.

Annexe 7. Combinaisons bactérie-antibiotiques pour évaluer l'antibiorésistance selon l'ECDC [37]

Espèce bactérienne	Famille d'antibiotiques	Antibiotiques
<i>Escherichia coli</i>	Aminopénicillines	Ampicilline ou amoxicilline
	C3G	Céfotaxime, ceftriaxone or ceftazidime
	Carbapénèmes	Imipénème ou méropénème
	Fluoroquinolones	Ciprofloxacine, lévofloxacine ou ofloxacine
	Aminoglycosides	Gentamicine, tobramycine ou nétilimicine [@]
<i>Klebsiella pneumoniae</i>	C3G	Céfotaxime, ceftriaxone or ceftazidime
	Carbapénèmes	Imipénème ou méropénème
	Fluoroquinolones	Ciprofloxacine, lévofloxacine ou ofloxacine
	Aminoglycosides	Gentamicine, tobramycine ou nétilimicine [@]
<i>Pseudomonas aeruginosa</i>	Pipéracilline-tazobactam	Pipéracilline-tazobactam
	Ceftazidime	Ceftazidime
	Carbapénèmes	Imipénème ou méropénème
	Fluoroquinolones	Ciprofloxacine ou lévofloxacine
	Aminoglycosides	Gentamicine, tobramycine ou nétilimicine [@]
<i>Streptococcus pneumoniae</i>	Pénicilline	Oxacilline ou pénicilline*
	Macrolides	Clarithromycine, érythromycine, azithromycine [@]
	Fluoroquinolones	Lévofloxacine ou moxifloxacine**
	C3G	Céfotaxime ou ceftriaxone
<i>Staphylococcus aureus</i>	SARM	Céfoxitine ou oxacilline
	Rifampicine	Rifampicine
	Fluoroquinolones	Lévofloxacine, ofloxacine ou ciprofloxacine [§]
<i>Enterococcus faecalis</i>	Gentamycine	Gentamycine (haut niveau)
	Vancomycine	Vancomycine

Abréviations : C3G : céphalosporines de troisième génération ; SARM : *S. aureus* résistant à la méticilline.

[@]Dans la base BactHub, les résultats des tests de sensibilité à la nétilimicine et à la clarithromycine ne sont pas disponibles et n'ont donc pas été considérés. Seule la résistance à la tobramycine a été considérée pour déterminer la résistance de *P. aeruginosa* aux aminoglycosides.

Règles ECDC : *La priorité est donnée aux résultats de sensibilité à la pénicilline plutôt qu'à ceux de l'oxacilline.

**Les résultats de sensibilité à la norfloxacine sont acceptés comme marqueurs de sensibilité aux fluoroquinolones, mais la priorité est donnée aux résultats de sensibilité à la lévofloxacine et à la moxifloxacine. [§]Les résultats de sensibilité à la norfloxacine sont acceptés comme marqueurs de sensibilité aux fluoroquinolones, mais la priorité est donnée aux résultats de sensibilité à la lévofloxacine, à la ciprofloxacine, et/ou à l'ofloxacine.

Annexe 8. Caractéristiques des patients avec bactériémie par année et groupe, base BactHub, 2016-2019

A. Patients avec un épisode communautaire

	2016	2017	2018	2019
Patients, N (%)	2 944	3 005	3 056	3 107
Sexe masculin, N (%)	1 672 (56,8)	1 625 (54,1)	1 676 (54,8)	1 705 (54,9)
Age, années	67 [54 - 79]	68 [55 - 80]	68 [54 - 80]	68 [55 - 80]
18-35	206 (7)	229 (7,6)	248 (8,1)	270 (8,7)
35-50	392 (13,3)	366 (12,2)	370 (12,1)	368 (11,8)
50-65	758 (25,8)	746 (24,8)	740 (24,2)	758 (24,4)
65-80	915 (31,1)	930 (31,0)	981 (32,1)	997 (32,1)
>80	673 (22,9)	734 (24,4)	717 (23,5)	714 (23,0)
Score de Charlson*	2 [0 - 3]	2 [0 - 3]	2 [0 - 3]	2 [0 - 3]
0	742 (38,8)	1049 (37,9)	1193 (40,0)	1175 (39,7)
1-2	595 (31,1)	933 (33,7)	935 (31,3)	955 (32,2)
3-4	247 (12,9)	337 (12,2)	400 (13,4)	347 (11,7)
5-6	102 (5,3)	141 (5,1)	150 (5,0)	127 (4,3)
7-8	154 (8,1)	230 (8,3)	224 (7,5)	262 (8,9)
>8	71 (3,7)	81 (2,9)	82 (2,8)	96 (3,2)
Comorbidités*, N (%)				
Cancer	477 (25,0)	714 (25,8)	745 (25,0)	773 (26,1)
Tumeur solide métastatique	199 (10,4)	268 (9,7)	272 (9,1)	315 (10,6)
Insuffisance cardiaque	290 (15,2)	427 (15,4)	435 (14,6)	420 (14,2)
Diabète	393 (20,6)	571 (20,6)	593 (19,9)	545 (18,4)
Non compliqué	253 (13,2)	346 (12,5)	378 (12,7)	338 (11,4)
Complication chronique	140 (7,3)	225 (8,1)	215 (7,2)	207 (7,0)
Maladie vasculaire	248 (13,0)	333 (12,0)	348 (11,7)	332 (11,2)
Coronaire	32 (1,7)	49 (1,8)	41 (1,4)	63 (2,1)
Périphérique	120 (6,3)	154 (5,6)	180 (6,0)	143 (4,8)
Cérébrovasculaire	113 (5,9)	151 (5,5)	167 (5,6)	150 (5,1)
Maladie rénale	270 (14,1)	361 (13,0)	366 (12,3)	340 (11,5)
Maladie hépatique	174 (9,1)	265 (9,6)	292 (9,8)	273 (9,2)
Légère	113 (5,9)	182 (6,6)	191 (6,4)	173 (5,8)
Modérée / sévère	61 (3,2)	83 (3,0)	101 (3,4)	100 (3,4)
Maladie pulmonaire chronique	154 (8,1)	189 (6,8)	193 (6,5)	180 (6,1)
Démence	101 (5,3)	136 (4,9)	163 (5,5)	132 (4,5)
Paralysie (hémiplégie / paraplégie)	53 (2,8)	76 (2,7)	81 (2,7)	86 (2,9)
Maladie systémique	25 (1,3)	42 (1,5)	42 (1,4)	45 (1,5)
Mortalité, N (%)				
Intra-hospitalière	424 (14,4)	426 (14,2)	376 (12,3)	410 (13,2)
Avec passage en soins intensifs	273 / 963 (28,4)	267 / 938 (28,5)	228 / 926 (24,6)	225 / 931 (24,2)
Sans passage en soins intensifs	151 / 1981 (7,6)	159 / 2067 (7,7)	148 / 2130 (7,0)	185 / 2176 (8,5)
À 30 jours	451 (15,3)	461 (15,3)	414 (13,6)	439 (14,1)
Avec passage en soins intensifs	262 / 963 (27,2)	254 / 938 (27,1)	226 / 926 (24,4)	209 / 931 (22,5)
Sans passage en soins intensifs	189 / 1981 (9,5)	207 / 2067 (10,0)	188 / 2130 (8,8)	230 / 2176 (10,6)
À 90 jours	588 (20,0)	627 (20,9)	574 (18,8)	587 (18,9)
Avec passage en soins intensifs	313 / 963 (32,5)	307 / 938 (32,7)	282 / 926 (30,5)	252 / 931 (27,1)
Sans passage en soins intensifs	275 / 1981 (13,9)	320 / 2067 (15,5)	292 / 2130 (13,7)	335 / 2176 (15,4)

B. Patients avec un épisode nosocomial

	2016	2017	2018	2019
Patients, N (%)	4 271	4 220	4 265	4 360
Sexe masculin, N (%)	2 648 (62,0)	2 619 (62,1)	2 618 (61,4)	2 691 (61,7)
Age, années	64 [52 - 74]	64 [51 - 74]	64 [52 - 74]	64 [52 - 74]
18-35	343 (8,0)	401 (9,5)	366 (8,6)	363 (8,3)
35-50	641 (15,0)	602 (14,3)	594 (13,9)	616 (14,1)
50-65	1 357 (31,8)	1 324 (31,4)	1 308 (30,7)	1 323 (30,3)
65-80	1 376 (32,2)	1 358 (32,2)	1 416 (33,2)	1 457 (33,4)
>80	554 (13,0)	535 (12,7)	581 (13,6)	601 (13,8)
Score de Charlson*	2 [1 - 5]	2 [1 - 6]	2 [1 - 4]	2 [1 - 4]
0	466 (19,4)	871 (22,8)	942 (22,2)	982 (23,2)
1-2	847 (35,3)	1 271 (33,2)	1 565 (36,9)	1 525 (36,0)
3-4	422 (17,6)	612 (16,0)	685 (16,1)	685 (16,2)
5-6	176 (7,3)	280 (7,3)	283 (6,7)	320 (7,6)
7-8	340 (14,2)	566 (14,8)	539 (12,7)	503 (11,9)
>8	148 (6,2)	226 (5,9)	229 (5,4)	219 (5,2)
Comorbidités*, N (%)				
Cancer	1 060 (44,2)	1 640 (42,9)	1 783 (42,0)	1 781 (42,1)
Tumeur solide métastatique	405 (16,9)	667 (17,4)	654 (15,4)	637 (15,0)
Insuffisance cardiaque	466 (19,4)	701 (18,3)	871 (20,5)	777 (18,4)
Diabète	519 (21,6)	731 (19,1)	866 (20,4)	824 (19,5)
Non compliqué	312 (13,0)	446 (11,7)	530 (12,5)	476 (11,2)
Complication chronique	207 (8,6)	285 (7,5)	336 (7,9)	348 (8,2)
Maladie vasculaire	430 (17,9)	718 (18,8)	794 (18,7)	774 (18,3)
Coronaire	62 (2,6)	123 (3,2)	167 (3,9)	144 (3,4)
Périphérique	218 (9,1)	347 (9,1)	352 (8,3)	320 (7,6)
Cérébrovasculaire	194 (8,1)	353 (9,2)	387 (9,1)	413 (9,8)
Maladie rénale	416 (17,3)	562 (14,7)	591 (13,9)	583 (13,8)
Maladie hépatique	310 (12,9)	493 (12,9)	482 (11,4)	472 (11,2)
Légère	161 (6,7)	258 (6,7)	267 (6,3)	259 (6,1)
Modérée / sévère	149 (6,2)	235 (6,1)	215 (5,1)	213 (5,0)
Maladie pulmonaire chronique	215 (9,0)	296 (7,7)	350 (8,3)	316 (7,5)
Démence	124 (5,2)	127 (3,3)	141 (3,3)	170 (4,0)
Paralysie (hémiplégie / paraplégie)	138 (5,8)	253 (6,6)	260 (6,1)	290 (6,9)
Maladie systémique	34 (1,4)	60 (1,6)	78 (1,8)	68 (1,6)
Mortalité, N (%)				
Intra-hospitalière	885 (20,7)	805 (19,1)	833 (19,5)	779 (17,9)
Avec passage en soins intensifs	564 / 1888 (29,9)	488 / 1860 (26,2)	543 / 1912 (28,4)	515 / 1955 (26,3)
Sans passage en soins intensifs	321 / 2383 (13,5)	317 / 2360 (13,4)	290 / 2353 (12,3)	264 / 2405 (11,0)
À 30 jours	786 (18,4)	745 (17,7)	769 (18,0)	728 (16,7)
Avec passage en soins intensifs	475 / 1888 (25,2)	414 / 1860 (22,3)	463 / 1912 (24,2)	436 / 1955 (22,3)
Sans passage en soins intensifs	311 / 2383 (13,1)	331 / 2360 (14,0)	306 / 2353 (13,0)	292 / 2405 (12,1)
À 90 jours	1 136 (26,6)	1 101 (26,1)	1 125 (26,4)	1 089 (25,0)
Avec passage en soins intensifs	635 / 1888 (33,6)	569 / 1860 (30,6)	622 / 1912 (32,5)	585 / 1955 (29,9)
Sans passage en soins intensifs	501 / 2383 (21,0)	532 / 2360 (22,5)	503 / 2353 (21,4)	504 / 2405 (21,0)

* Données disponibles pour 10 628 et 14 702 patients avec épisodes CO et HA.

Annexe 9. Taux de résistance aux antibiotiques (en pourcentage) des isolats bactériens dans les épisodes communautaires et nosocomiaux, globalement et par année, base BactHub, 2016-2019

	Global	2016	2017	2018	2019	Tendance
<i>Escherichia coli</i>	N = 6 983	N = 1 711	N = 1 762	N = 1 766	N = 1 744	
Aminopénicillines	64,94	65,98	67,03	62,26	64,52	
C3G	14,95	16,95	14,73	14,59	13,57	<
Carbapénèmes	0,25	0,00	0,23	0,29	0,46	>
Fluoroquinolones	22,42	27,27	20,73	20,79	20,99	<
Aminoglycosides	11,60	13,83	11,59	10,54	10,49	<
<i>Staphylococcus aureus</i>	N = 4 185	N = 1 022	N = 1 045	N = 1 063	N = 1 055	
SARM	12,98	13,99	13,73	11,89	12,27	
Rifampicine	1,56	1,41	1,83	1,20	1,80	
Fluoroquinolones	11,42	12,55	11,65	9,35	12,16	
<i>Klebsiella pneumoniae</i>	N = 2 254	N = 556	N = 531	N = 575	N = 592	
C3G	31,70	31,59	28,19	33,39	33,33	
Carbapénèmes	1,39	2,0	1,14	1,40	1,03	
Fluoroquinolones	29,30	30,69	27,05	30,40	28,94	
Aminoglycosides	23,19	24,91	21,67	25,13	21,03	
<i>Staphylococcus epidermidis</i>	N = 1 757	N = 430	N = 464	N = 456	N = 407	
Métiline	81,5	81,63	79,74	83,55	81,08	
Rifampicine	14,04	9,53	12,20	18,42	15,97	>
Fluoroquinolones	67,52	70,70	65,44	68,13	65,84	
<i>Pseudomonas aeruginosa</i>	N = 1 772	N = 503	N = 426	N = 395	N = 448	
Pipéracilline-tazobactam	15,37	17,38	13,04	17,74	12,63	
Ceftazidime	10,47	11,02	9,69	11,68	9,52	
Carbapénèmes	19,86	22,55	18,44	18,27	19,59	
Fluoroquinolones	12,06	16,4	11,58	11,45	8,14	<
Aminoglycosides*	6,48	9,18	5,21	7,63	3,61	<
<i>Streptococcus pneumoniae</i>	N = 639	N = 159	N = 137	N = 169	N = 174	
Pénicilline**	21,13	20	13,33	24,06	25,6	
Macrolides	22,00	22,15	20,16	22,84	22,49	
Fluoroquinolones	4,4	11,4	4,1	1,4	2,0	<
C3G	2,09	2,21	0,85	1,46	3,68	
<i>Enterococcus faecalis</i>	N = 1 517	N = 365	N = 384	N = 375	N = 393	
Gentamycine (haut niveau)	9,27	11,52	8,79	8,86	7,95	
Vancomycine	0,34	0,84	0,00	0,55	0,00	

Abréviations : C3G : céphalosporines de troisième génération ; SARM : *S. aureus* résistant à la métiline. La significativité statistique des tendances a été évaluée par un test de χ^2 pour la tendance, et une valeur p de <0,05 a été considérée comme significative. Seuls les résultats statistiquement significatifs sont présentés. Pour tous les antibiotiques testés, sauf la gentamycine pour *E. faecalis*, les souches intermédiaires étaient considérées résistantes. * Résistance à la tobramycine. ** La priorité a été donnée aux résultats de sensibilité à la pénicilline plutôt qu'à ceux de l'oxacilline.

Annexe 10. Taux de résistance aux antibiotiques (en pourcentage) des isolats bactériens et tendance dans la base BactHub et dans les données de l'ECDC

	BactHub (2016 - 2019)	Tendance 2016 - 2019	ECDC France (2016 - 2019)	Tendance 2015 - 2019	Tous les pays de l'ECDC (2016 - 2019)	Tendance 2015 - 2019
<i>E. coli</i>						
Aminopénicillines	64,94		57,2% - 54,5%	<	59,0% - 57,1%	<
C3G	14,95	<	11,2% - 8,8%	<	14,9% - 15,1%	>
Carbapénèmes	0,25	>	0% - 0%		0,1% - 0,3%	>
Fluoroquinolones	22,42	<	16,7% - 16,0%	<	25,2% - 23,8%	<
Aminoglycosides	11,60	<	7,9% - 7,0%	<	11,6% - 10,8%	<
<i>S. aureus</i>						
SARM	12,98		13,8% - 11,6%	<	17,7% - 15,5%	<
<i>K. pneumoniae</i>						
C3G	31,70		28,9% - 30,2%		31,4% - 31,3%	
Carbapénèmes	1,39		0,4% - 1,0%	>	7,4% - 7,9%	>
Fluoroquinolones	29,30		27,7% - 30,9%		30,3% - 31,2%	>
Aminoglycosides	23,19		26,2% - 23,4%	<	24,4% - 22,3%	<
<i>P. aeruginosa</i>						
Pip-tazobactam	15,37		16,0% - 16,7%	>	17,5% - 16,9%	<
Ceftazidime	10,47		11,3% - 11,5%		14,4% - 14,3%	<
Carbapénèmes	19,86		15,6% - 12,7%	<	18,2% - 16,5%	<
Fluoroquinolones	12,06	<	13,6% - 13,7%	<	18,8% - 18,9%	<
Aminoglycosides*	6,48	<	10,7% - 7,8%	<	14,0% - 11,5%	<
<i>S. pneumoniae</i>						
Pénicilline**	21,13		25,3% - 25,3%		13,1% - 12,1%	<
Macrolides	22,00		22,9% - 19,4%	<	16,6% - 14,5%	<
<i>E. faecalis</i>						
Gentamycine [§]	9,27		15,0% - 12,0%	<	31,8% - 26,6%	<

Abréviations : C3G : céphalosporines de troisième génération ; Pip-tazobactam : pipéracilline-tazobactam ; SARM : *S. aureus* résistant à la méticilline. La significativité statistique des tendances a été évaluée par un test de χ^2 pour la tendance, et une valeur p de <0,05 a été considérée comme significative. Seuls les résultats statistiquement significatifs sont présentés. Pour tous les antibiotiques testés, sauf la gentamycine pour *E. faecalis*, les souches intermédiaires étaient considérées résistantes. * La résistance à la tobramycine de *P. aeruginosa* dans la base BactHub a été comparée à la résistance à la tobramycine / gentamycine / nétilmicine de *P. aeruginosa* dans les données de l'ECDC. ** La priorité a été donnée aux résultats de sensibilité à la pénicilline plutôt qu'à ceux de l'oxacilline. [§] Résistance de haut niveau à la gentamycine.

Annexe 11. Odds ratio ajustés et intervalles de confiance à 95% du lien entre chaque couple bactérie-résistance et la récurrence de bactériémie à 1 an dans le modèle multivarié final, selon la bactérie-résistance prise en référence*

Bactérie et résistance			
SAMS	1	0,87 [0,6 - 1,26]	0,70 [0,44 - 1,14]
SARM	0,77 [0,23 - 1,99]	0,67 [0,2 - 1,71]	0,54 [0,16 - 1,45]
<i>E. coli</i> non C3G-R	1,15 [0,79 - 1,68]	1	0,80 [0,54 - 1,23]
<i>E. coli</i> C3G-R	2,48 [1,53 - 4,01]	2,16 [1,44 - 3,2]	1,74 [1,05 - 2,91]
<i>Klebsiella spp</i> non C3G-R	1,43 [0,88 - 2,28]	1,24 [0,82 - 1,85]	1
<i>Klebsiella spp</i> C3G-R	4,70 [2,52 - 8,58]	4,09 [2,29 - 7,08]	3,30 [1,72 - 6,21]

*Les résultats présentés sont ajustés sur l'ensemble des variables décrites dans le Tableau 7.2

Annexe 12. Analyses descriptives et analyses multivariées des facteurs de risque de récurrence à 1 an suite à un épisode incident communautaire monomicrobien à *S. aureus*, *E. coli* ou *Klebsiella spp*, dans le sous-échantillon avec information sur l'adéquation du traitement empirique, base BactHub, 2017-2019

	Description de la population, N (%)		Analyse multivariée	
	Avec récurrence N = 117	Sans récurrence N = 1175	OR [IC95]	p-value
Patients				
Sexe				0,40
Homme	67 (57,3)	522 (44,4)	1	
Femme	50 (42,7)	653 (55,6)	0,84 [0,55 - 1,27]	
Age, années				0,009
18-35	1 (0,9)	80 (6,8)	0,15 [0,01 - 0,84]	
35-50	12 (10,3)	130 (11,1)	1	
50-65	39 (33,3)	254 (21,6)	1,57 [0,79 - 3,36]	
65-80	44 (37,6)	392 (33,4)	1,08 [0,54 - 2,27]	
>80	21 (17,9)	319 (27,1)	0,74 [0,34 - 1,66]	
Comorbidités*				
Cancer	40 (34,5)	218 (19)	1,9 [1,2 - 3]	0,007
Maladie rénale	25 (21,6)	171 (14,9)	1,66 [0,98 - 2,75]	0,06
Maladie hépatique	16 (13,8)	82 (7,2)	1,7 [0,88 - 3,13]	0,11
Épisodes incidents				
Site d'infection primaire*				0,009
Bactériémie primaire	25 (21,6)	127 (11,1)	3,26 [1,76 - 6,03]	
Bactériémie secondaire				
Infections multiples	27 (23,3)	262 (22,9)	1,76 [0,97 - 3,17]	
Infection respiratoire basse	6 (5,2)	47 (4,1)	2,32 [0,79 - 5,9]	
Infection urinaire	30 (25,9)	481 (42)	1	
Infection digestive	11 (9,5)	104 (9,1)	1,72 [0,77 - 3,63]	
Infection sur matériel	12 (10,3)	52 (4,5)	3,26 [1,39 - 7,39]	
Autres	5 (4,3)	73 (6,4)	1,75 [0,54 - 4,78]	
Bactérie et résistance				0,002
SAMS	17 (14,5)	205 (17,4)	1	
SARM	2 (1,7)	24 (2)	0,86 [0,13 - 3,47]	
<i>E. coli</i> non C3G-R	54 (46,2)	705 (60)	1,43 [0,77 - 2,77]	
<i>E. coli</i> C3G-R	20 (17,1)	113 (9,6)	2,78 [1,21 - 6,4]	
<i>Klebsiella spp</i> non C3G-R	13 (11,1)	105 (8,9)	1,42 [0,61 - 3,23]	
<i>Klebsiella spp</i> C3G-R	11 (9,4)	23 (2)	9,14 [3,12 - 26,62]	
Traitement empirique				0,38
Adapté	95 (81,2)	1072 (91,2)	1	
Inadapté	22 (18,8)	103 (8,8)	1,35 [0,69 - 2,57]	

Abréviations : C3G-R : résistant aux céphalosporines de 3^{ème} génération ; Q1 : premier quartile ; Q3 : troisième quartile ; SAMS / SARM : *S. aureus* sensible / résistant à la méticilline. * Données manquantes : 1 épisode avec récurrence, 29 épisodes sans récurrence.

Annexe 13. Études cas-témoins évaluant l'exposition aux antibiotiques comme potentiel facteur de risque de résistance dans les bactériémies et les infections urinaires communautaires

Premier auteur, année	Groupe étudié	Paire bactérie-résistance étudiée
Bactériémies		
Rodríguez-Baño, 2010 [209]	C	<i>E. coli</i> BLSE
Hsieh, 2010 [211]	C	<i>E. coli</i> BLSE
Park, 2011 [161]	C	<i>E. coli</i> BLSE
Park, 2014 [162]	C	<i>E. coli</i> BLSE
Kim, 2019 [163]	C	<i>E. coli</i> ou <i>Klebsiella</i> spp. BLSE
Lee, 2017 [164]	C	Entérobactéries BLSE
Zahar, 2017 [210]	C	Entérobactéries BLSE
Lin, 2019 [250]	C	<i>E. coli</i> C3G-R
Lee, 2014 [167]	C	<i>E. coli</i> , <i>K. pneumoniae</i> C3G-R
Sung, 2012 [168]	C vraies (en réalité C*)	<i>E. coli</i> , <i>Klebsiella</i> , <i>Enterobacter</i> , <i>Citrobacter</i> C3G-R
Gottesman, 2018 [169]	C vraies	<i>E. coli</i> BLSE
Pedersen, 1999 [170]	C vraies	<i>E. Coli</i> R à l'ampicilline, aux sulfamides, au triméthoprim
Ortega, 2016 [171]	C vraies	<i>E. coli</i> ou <i>Klebsiella</i> spp. C3G-R
Wi, 2018 [172]	C	SARM
Lim, 2014 [206]	C	SARM
Infections urinaires		
Fan, 2014 [174]	C	<i>E. coli</i> BLSE
Pérez Heras, 2017 [175]	C	<i>E. coli</i> BLSE
Søgaard, 2017 [212]	C	<i>E. coli</i> BLSE
Tüzün, 2019 [177]	C	<i>E. coli</i> BLSE
Boix-Palop, 2017 [178]	C	<i>K. pneumoniae</i> BLSE
Kung, 2015 [179]	C	Entérobactéries BLSE
Calbo, 2006 [180]	C (en réalité C vraies**)	<i>E. coli</i> BLSE
Zhu, 2019 [181]	C vraies (en réalité C)	<i>E. coli</i> BLSE
Topaloglu, 2010 [182]	C vraies (en réalité C)	<i>E. coli</i> ou <i>Klebsiella</i> spp. BLSE
Yilmaz, 2008 [183]	C/C vraies	<i>E. coli</i> ou <i>K. pneumoniae</i> BLSE
Colodner, 2004 [184]	Non hospitalisées	<i>E. coli</i> ou <i>Klebsiella</i> spp. BLSE
Albaramki, 2019 [185]	C vraies	<i>E. coli</i> ou <i>Klebsiella</i> spp. BLSE
Guzmán, 2019 [186]	C vraies	<i>E. coli</i> BLSE
Koksal, 2019 [187]	C vraies	<i>E. coli</i> ou <i>Klebsiella</i> spp. BLSE
Azap, 2010 [188]	C vraies	<i>E. coli</i> BLSE
Rodríguez-Baño, 2008 [189]	C vraies	<i>E. coli</i> BLSE
Castillo-Tokumori, 2017 [190]	C vraies	<i>E. coli</i> BLSE
Megged, 2014 [191]	C vraies	Entérobactéries BLSE
Chervet, 2018 [192]	C vraies (en réalité C)	Entérobactéries BLSE
Jacmel, 2017 [193]	C vraies (en réalité C)	Entérobactéries BLSE
Brosh-Nissimov, 2019 [194]	C vraies (en réalité C)	Entérobactéries BLSE
Colodner, 2008 [195]	C vraies (en réalité C)	<i>E. coli</i> R aux quinolones
Seija, 2014 [196]	C vraies (en réalité C)	<i>E. coli</i> R aux fluoroquinolones
Killgore, 2004 [197]	C vraies (en réalité C)	<i>E. coli</i> R à la ciprofloxacine
Chaniotaki, 2004 [198]	C vraies (en réalité C)	<i>E. coli</i> R aux quinolones
Dromigny, 2005 [199]	C vraies (en réalité C)	<i>E. coli</i> R à l'ampicilline, à l'amoxicilline, à l'acide nalidixique, aux fluoroquinolones, au cotrimoxazole
Hillier, 2007 [200]	C vraies (en réalité C)	Toutes bactéries, R à l'ampicilline, au triméthoprim

Anesi, 2016 [201]	C	Entérobactéries C3G-R
Lee, 2015 [202]	C	Entérobactéries avec AmpC dérégulée
Smithson, 2012 [203]	C vraies	<i>E. coli</i> R aux quinolones
Bosch-Nicolau, 2017 [204]	C vraies (en réalité C)	Toutes bactéries, R à l'amoxicilline-clavulanate, au céfuroxime, au céfotaxime, à la ciprofloxacine
Filiatrault et al., 2012 [205]	C vraies	Toutes bactéries, R à la ciprofloxacine, au cotrimoxazole

Abréviations : BLSE : β -lactamases à spectre étendu ; C = communautaires ; C3G-R : résistant aux céphalosporines de 3^{ème} génération ; IU : infection urinaire ; R : résistant. *Études où les infections étudiées sont désignées comme étant communautaires vraies, mais dont la définition est plutôt proche des infections communautaires, donc inclut de potentielles expositions aux soins. **Études où les infections étudiées sont désignées comme étant communautaires, mais dont la définition est plutôt proche des infections communautaires vraies, donc exclut des critères d'exposition aux soins.

Annexe 14. Codes CIM-10 utilisés pour sélectionner les témoins

Codes d'infection aigüe ou chronique à bactérie ou à mycobactérie	A00, A01, A02, A03, A04, A05, A15, A16, A17, A18, A19, A20, A21, A22, A23, A24, A25, A26, A27, A28, A30, A31, A32, A33, A34, A35, A36, A37, A38, A39, A40, A41, A42, A43, A44, A46, A48, A49, A50, A51, A52, A53, A54, A55, A56, A57, A58, A65, A66, A67, A68, A69, A70, A71, A74, A75, A77, A78, A79, B95, B96, D733, G00, G01, G03.9, G04.2, G05.0, G06, G07, H04.3, H04.4, H10.0, H10.3, H10.4, H10.5, H15, H15.0, H151, H16, H16.0, H16.1, H16.2, H16.3, H16.4, H16.8, H16.9, H30, H30.0, H30.1, H30.2, H30.8, H30.9, H44.0, H60, H61.0, H62.0, H66, H67.0, H68, H70, H71, H73, H75, I30.1, I30.9, I32.0, I33, I40.0, I41.0, I70.01, I70.21, I70.81, I70.91, I98.0, J01, J02, J03, J04, J05, J06, J13, J14, J15, J16, J17.0, J18, J20.0, J20.1, J20.2, J20.8, J22, J32, J34.0, J35.0, J36, J37, J39.0, J39.1, J40, J41, J42, J85, J86, J90, J98.5, K04.4, K04.5, K04.6, K04.7, K05.0, K05.1, K05.2, K05.3, K10.2, K10.3, K11.3, K12.2, K23.0, K29.1, K29.5, K29.6, K29.7, K29.8, K29.9, K35, K36, K37, K40.1, K40.4, K41.1, K41.4, K42.1, K43.1, K43.4, K43.7, K44.1, K45.1, K57.0, K57.2, K57.4, K57.8, K61, K63.0, K65, K67, K75.0, K80.0, K80.1, K80.3, K80.4, K81, K83.0, K93.0, L00, L01, L02, L03, L04, L05.0, L08, L30.3, L73.2, L88, M00.0, M00.1, M00.2, M00.8, M00.9, M01.3, M46.2, M46.3, M46.4, M46.5, M49.1, M49.2, M49.3, M60.0, M65.0, M71.0, M71.1, M72.6, M86, N08.0, N10, N11.0, N11.1, N12, N13.6, N15.1, N16.0, N29.0, N29.1, N30.0, N30.2, N30.3, N30.8, N30.9, N33.0, N34, N39.0, N41, N45, N49, N51, N61, N70, N71, N72, N73, N74, N75, N76, N77, O23, O41.1, O85, O86, O88.3, O91, O98.0, O98.1, O98.2, O98.8, O98.9, P00.2, P23.1, P23.2, P23.3, P23.4, P23.5, P23.6, P23.8, P36, P37.0, P37.2, P37.8, P39, P78.1, R02, R57.2, R65.0, R65.1, T31, T79.3, T80.2, T81.4, T82.6, T82.7, T83.5, T83.6, T84.5, T84.6, T84.7, T85.7, T87.4, T88.0, U80, U81, U82, U83, U84, U85, U88, U89, Z76.800
--	--
