



Fiche Résumé de thèse

La fiche résumé est à envoyer par mail à l'adresse dgs.prixdethese@sante.gouv.fr avec l'ensemble des pièces du dossier de candidature jusqu'au 15 janvier 2023 inclus.

Informations sur la thèse

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

Problématique : Antibiorésistance, Entrepôts de données de santé hospitaliers

Discipline : Santé publique - Épidémiologie



Résumé de la thèse

La résistance des bactéries aux antibiotiques, désignée sous l'appellation antibiorésistance, est un enjeu majeur de santé publique. En effet, de plus en plus de bactéries deviennent résistantes aux antibiotiques, et ces bactéries se diffusent dans la population. Par conséquent, les antibiotiques deviennent moins efficaces pour traiter les infections bactériennes. En 2019, près de 5 millions de décès dans le Monde seraient associés à l'antibiorésistance, soit près de 9% des décès globaux.

Pour contrôler l'augmentation des bactéries résistantes, le gouvernement français a mis en place des stratégies de santé publique dans le cadre de la feuille de route antibiorésistance en 2016, et de la Stratégie nationale de prévention des infections et de l'antibiorésistance 2022-2025. Ces stratégies sont guidées par les données scientifiques. Cependant, de nombreuses questions scientifiques restent en suspens, pour mieux comprendre les conséquences de l'antibiorésistance sur la santé, et préciser les facteurs qui participent à l'apparition et à la diffusion de l'antibiorésistance. Répondre à ces questions complexes nécessite des données à la fois sur les bactéries résistantes aux antibiotiques, et sur les patients qui sont infectés par ces bactéries. La France dispose de nombreuses données de qualité de surveillance de l'antibiorésistance, mais qui se focalisent sur les bactéries, sans information sur les patients infectés. Une autre source de données intéressante sont les données hospitalières, qui contiennent à la fois des données sur des patients hospitalisés infectés par une bactérie, et sur les bactéries en cause et leurs résistances aux antibiotiques.

Ainsi, l'objectif de ma thèse était d'évaluer l'apport des données hospitalières pour l'étude des infections bactériennes résistantes aux antibiotiques. Je me suis focalisée sur les infections bactériennes dans le sang (bactériémies), qui sont des infections sévères couramment rencontrées par les soignants, en ville ou à l'hôpital. J'ai analysé des données issues de l'Entrepôt de Données de Santé de l'Assistance Publique – Hôpitaux de Paris (AP-HP), qui regroupe les données informatisées de tous les patients hospitalisés dans un hôpital de l'AP-HP. L'AP-HP est le plus grand regroupement d'hôpitaux d'Europe, comptant 39 hôpitaux situés principalement dans la région parisienne, et représentant plus de 10% des hospitalisations en France.

1. Construction de la base BactHub et caractérisation des patients

Dans un premier temps, j'ai construit la base BactHub, qui contient les données de plus de 30 000 patients hospitalisés avec une bactériémie entre 2016 et 2019. Il s'agit de la première base en France, et l'une des rares à l'échelle internationale, avec des données aussi détaillées pour un grand nombre de patients avec une infection bactérienne aiguë, et qui inclut des données à la fois sur les patients et sur leurs bactéries et leurs résistances aux antibiotiques. Construire cette base a nécessité plus d'un an et demi de travail informatique et la création de nombreux algorithmes informatiques complexes.

J'ai ensuite spécifié, dans une publication scientifique dans la revue *Clinical Epidemiology*, les principales caractéristiques des patients (âge, sexe, antécédents, mortalité, etc.), de leur séjour (durée, etc.), et des bactéries (nom, résistance aux antibiotiques) identifiées dans leur sang. Mes résultats montrent de forts taux de mortalité, atteignant plus de 20% à trois mois de l'infection, mais qui sont plus faibles dans le cas d'une bactérie acquise en ville plutôt qu'à l'hôpital. Les bactéries associées aux plus forts taux de mortalité étaient *Pseudomonas aeruginosa* et *Enterococcus faecium* (~30% et ~40%, respectivement). Par ailleurs, toutes les bactéries présentaient des taux importants de résistance aux antibiotiques. Par exemple, j'ai montré que ~20-35% des bactéries de type *Klebsiella pneumoniae* et ~15% des bactéries de type *Escherichia coli* identifiées dans le sang présentaient une résistance aux céphalosporines de troisième génération. Cette résistance s'associe à une augmentation du risque de mortalité, et inquiète à l'échelle nationale et internationale car elle diffuse dans la population. Je décris de nombreux autres taux de résistance, comme la résistance des bactéries de type *Staphylococcus aureus* à la méthicilline, qui est de 10-15 %.

2. Utilisation des algorithmes de traitement automatique du langage naturel

Le traitement automatique du langage naturel est un des champs de recherche les plus actifs en intelligence artificielle et en science des données, qui vise à apprendre à une machine à comprendre



le langage humain. Appliqué aux comptes-rendus médicaux, un algorithme de traitement automatique du langage naturel pourrait récupérer des informations d'intérêt de manière automatique dans le texte. Dans le cadre de ma thèse, nous avons développé avec le Pr Xavier Tannier un algorithme afin d'identifier dans les comptes-rendus médicaux d'où vient la bactérie trouvée dans le sang. En effet, avant d'être dans le sang, une bactérie peut venir d'une autre infection comme une infection urinaire, digestive, cutanée, etc. L'évaluation des performances de l'algorithme montre que quand il identifie une information, elle est souvent correcte. Par contre, l'algorithme ne peut rien identifier dans un grand nombre de cas, car plus de 25% des compte-rendus des patients ne contenaient pas d'information sur leur infection. Ainsi, si le développement du traitement automatique du langage naturel en santé est prometteur, des actions doivent être menées pour que les comptes-rendus médicaux soient remplis de manière plus exhaustive et qu'ils soient plus homogènes.

3. Impact de l'antibiorésistance sur le risque de refaire une infection

A ce jour, on connaît mal l'impact de l'antibiorésistance sur le risque de refaire une infection. Dans une étude portant sur plus de 3 500 patients avec une bactériémie acquise en ville aux bactéries *Klebsiella spp.*, *E. coli*, ou *S. aureus*, j'ai montré pour la première fois que l'antibiorésistance est un facteur de risque important de refaire une bactériémie dans l'année, pour les bactéries *Klebsiella spp.* ou *E. coli* avec une résistance aux céphalosporines de troisième génération. Cette étude a également identifié d'autres facteurs de risque de refaire une infection, qui sont le fait d'avoir un cancer, une maladie hépatique, une maladie rénale, et le fait que la bactériémie soit associée à une infection digestive ou à une infection liée à un dispositif médical. Ces résultats révèlent un nouvel impact de la diffusion de la résistance aux céphalosporines de troisième génération dans la population, et permettront de guider les stratégies de santé publique liées à l'antibiorésistance et la prise en charge des patients. L'article scientifique de cette étude est en cours d'évaluation par une revue internationale.

4. Lien entre prise d'antibiotiques et antibiorésistance

Ma thèse s'est également intéressée au lien entre antibiotiques et antibiorésistance. Si de nombreuses études concernent le milieu hospitalier, on connaît mal à ce jour le risque des individus dans la population de présenter une infection bactérienne antibiorésistante après avoir pris des antibiotiques. Dans une publication dans la revue *Antibiotics*, j'ai discuté les études publiées à ce sujet en termes de méthodologie et de résultats, et fait une proposition pour standardiser les futures études. Par ailleurs, nous avons obtenu en juillet 2022 l'autorisation de la CNIL afin d'enrichir la base Bacthub avec des données issues de la base principale du Système National des Données de Santé. Ceci nous permettra de savoir, pour les patients hospitalisés avec une bactériémie acquise en ville, quels sont les antibiotiques qu'ils ont pris avant leur infection, et donc d'évaluer précisément le lien entre prise d'antibiotiques et risque de bactériémie résistante aux antibiotiques.

5. Les entrepôts de données de santé hospitaliers pour guider l'action publique

A travers le cas d'usage antibiorésistance, cette thèse montre que la constitution croissante d'entrepôts de données de santé hospitaliers présente un potentiel important pour guider les stratégies de santé publique. Ces entrepôts donnent accès à des données variées issues de la pratique clinique hospitalière, et permettent de réaliser des études en incluant un nombre important de patients sans retour aux dossiers médicaux. Pour tirer le maximum de ces données, de nombreux défis doivent être relevés, notamment des défis techniques, légaux, éthiques, et humains. Ainsi, il est nécessaire de créer un cadre de confiance dans la population, qui favorise le partage des données dans le respect des exigences de sécurité et de confidentialité liées aux données de santé. Par ailleurs, il paraît indispensable d'accroître la formation des soignants à l'informatique et aux bases de données, et de mieux former les experts en données aux enjeux de santé. Enfin, il est nécessaire de maximiser la capacité des entrepôts à fonctionner avec d'autres bases de données, et de s'approcher de référentiels communs de structuration et de stockage des données. Ainsi, les données des entrepôts de données de santé pourraient être mises en lien avec d'autres bases de données, comme la base principale du Système National des Données de Santé, avec un intérêt majeur pour la recherche et pour guider l'action publique.