

# Santé numérique au Groupe Santé CHC : l'analyse de données et l'intelligence artificielle pour l'amélioration de la qualité des soins au sein d'un réseau hospitalier

Robin VERSCHUREN<sup>1</sup>, Philippe OLIVIER<sup>2</sup>, Guillaume VANHALST<sup>3</sup>, Christine THIRION<sup>4</sup>,  
Samy Amir AOUACHRIA<sup>5</sup>, Elliott BERTRAND<sup>6</sup>

<sup>1</sup> Effixis, [robin.verschuren@effixis.be](mailto:robin.verschuren@effixis.be)

<sup>2</sup> Groupe Santé CHC, [philippe.olivier@chc.be](mailto:philippe.olivier@chc.be)

<sup>3</sup> Effixis, [guillaume.vanhalst@effixis.be](mailto:guillaume.vanhalst@effixis.be)

<sup>4</sup> Groupe Santé CHC, [christine.thirion@chc.be](mailto:christine.thirion@chc.be)

<sup>5</sup> Groupe Santé CHC, [amir-samy.aouchria@chc.be](mailto:amir-samy.aouchria@chc.be)

<sup>6</sup> Effixis, [elliott.bertrand@effixis.ch](mailto:elliott.bertrand@effixis.ch)

## Résumé.

**Objectif :** L'utilisation d'Entrepôts de Données de Santé (EDS) est devenue cruciale pour l'amélioration des soins et la prise de décisions dans les hôpitaux modernes. Ce projet se concentre sur le déploiement d'un EDS au sein du réseau d'hôpitaux du Groupe Santé CHC.

**Méthodologie :** En employant une méthodologie centrée sur des cas d'usage, l'étude démontre comment l'EDS est un outil clé dans la prise de décisions et l'amélioration des soins aux patients. Les données, qui couvrent la période de 2008 à 2023, englobent des informations administratives et médicales relatives à 2,9 millions de séjours. La mise en place de l'interface entre l'EDS et ses utilisateurs s'est faite au travers d'un tableau de bord. Pour simplifier l'interaction entre l'utilisateur et les données, nous avons exploré expérimentalement l'intégration d'une intelligence artificielle générative, capable de répondre à des questions et générer des visualisations.

**Résultats :** Cette méthodologie a permis d'obtenir des résultats significatifs pour plusieurs cas d'usage effectivement utilisés, comme l'étude sur l'appendicectomie, des ménisectomies et l'examen approfondi de la gestion du diabète. L'IA générative a notamment permis d'évaluer la mortalité post-fracture de la hanche.

**Mots clés :** Réseau hospitalier, Entrepôts de Données de Santé (EDS), Benchmarking, Intelligence artificielle, Grand Modèle de Language.

## Introduction

En consolidant les fichiers de données sur une période s'étendant de 2008 à 2023, le groupe hospitalier a mesuré l'intérêt majeur d'un EDS qui repose sur une approche multi-sources et longitudinale dans le temps « long ». L'EDS permet de basculer d'une vision classiquement cloisonnée par année d'activité à une vision de type « populationnelle ».

Avec plus de 2,9 millions de séjours entre 2008 et 2023, le Groupe Santé CHC reconnaît l'importance de l'utilisation des EDS dans le traitement des données médicales. C'est notamment par le biais de ce projet de transformation numérique, que nous visons à améliorer la prise de décision clinique basée sur les données. L'institution, qui regroupe plusieurs sites hospitaliers, se concentre également sur la réalisation d'études comparatives entre les hôpitaux.

L'approche est basée sur des cas d'usage concrets et démontre l'impact potentiel de l'EDS sur la qualité des soins, via des indicateurs solides et actualisés, et sur l'orientation des décisions stratégiques pilotées par les

données. En combinant la technologie des données avec une interface homme-machine intuitive, l'analyse des données médicales devient rapide et accessible.

L'amélioration de la qualité des soins est également un objectif des autorités de tutelle, notamment avec les indicateurs de qualité communs aux hôpitaux bruxellois et wallons au travers de la plateforme PAQS. Cependant, la mise en place et l'évaluation de ces indicateurs sont souvent entravées par le manque d'outils adaptés et la présence de données non structurées. La mise à jour de ces indicateurs est au mieux « semi-manuelle » et pêche donc par manque de récurrence et donc d'actualisation. Il y a donc un intérêt manifeste pour un outil, qui automatise ces indicateurs et permet dès lors d'améliorer la mise en place d'une politique de qualité pilotée par ces mêmes indicateurs.

## **1. Contexte et travaux relatifs**

L'intérêt croissant pour les EDS découle du besoin de centraliser les données de santé, en particulier face à l'explosion des sources numériques disponibles [Haute Autorité de Santé, 2020, 2022]. Lancry (2012) met en évidence l'utilité des données pour optimiser l'allocation des ressources hospitalières. Ce regroupement de données permet également de les rendre exploitables pour la gestion administrative et opérationnelle, la recherche scientifique, ainsi que l'amélioration des soins prodigués aux patients [Hardy (2017)]. Cela permet de créer une correspondance entre les besoins variés des utilisateurs — qu'ils soient du personnel médical, administratif, ou autre — et des réponses concrètes, basées sur des données existantes, mais pas forcément récoltées à cet effet.

Ce projet s'inscrit dans le contexte et les défis inhérents aux données de santé, tels que la garantie de la qualité des données, l'interopérabilité entre divers systèmes et formats de données, ainsi que les enjeux de gouvernance et de transparence qui sont notamment soulignées dans les recherches de Griffier (2020). L'utilisation secondaire des données est une préoccupation essentielle.

Plus récemment, Wang (2023) a souligné l'intérêt d'utiliser les Grands Modèles de Langage (LLM - Large Language Models) pour prédire la requête SQL à partir d'une question sur une base de données, puis en exécutant cette requête sur la base de données.

L'implémentation d'un EDS au sein d'une structure hospitalière en réseau a permis d'en illustrer toute l'utilité à travers plusieurs cas concrets d'utilisation, sans aucune exhaustivité à ce stade, mais ouvrant des perspectives d'avenir sur la qualité des soins, le pilotage institutionnel « Data driven » et la recherche médicale. Il s'agit bien d'une première étape d'un chantier sur le long terme toujours en phase de recherche et développement.

## **2. Méthodologie**

Par une démarche itérative continuellement alimentée par les cas d'usages et les questionnements opérationnels, la pertinence d'un EDS se révèle progressivement en répondant parfois à des questionnements insolubles autrement.

### **2.1. Structure du Groupe Santé CHC**

Cette recherche est réalisée au sein du Groupe santé CHC, qui détient trois numéros d'agrément et forme un réseau de 5 hôpitaux situés dans la province de Liège en Belgique. Les cas d'utilisation sont intégrés

progressivement, en réponse aux demandes internes et aux besoins identifiés, ce qui favorise une approche dynamique et adaptative dans le développement de l'EDS.

Les analyses effectuées permettent notamment de comparer les différents sites hospitaliers. L'intérêt ici est de montrer que cet EDS peut être étendu à d'autres hôpitaux. En effet, tous les hôpitaux belges utilisent la même structure de données dans le Résumé Hospitalier Minimum (RHM), qui est définie au niveau fédéral. Cela signifie qu'une analyse comparative entre deux ou plusieurs sites hospitaliers pourrait être réalisée, tant que les données sont intégrées à l'EDS.

## **2.2. Données**

Les données sélectionnées pour le projet incluent des données administratives (provenant de la 3ème directive fédérale RHM) [SPF Santé publique, 2016] et médicales (5ème directive fédérale) de 2008 à 2023, des facturations de codes ATC A10 pour l'étude de l'insuline, et des listes des patients sous convention INAMI « diabète ». En outre, les données DRG du SPF Santé sont intégrées pour enrichir l'analyse. Lorsque ces données DRG n'ont pas encore été envoyées par le SPF Santé, des prédictions sont ajoutées afin que les analyses puissent être réalisées même pour des séjours récents. Nous disposons d'une base de données englobant approximativement 2,9 millions de séjours hospitaliers. Les données peuvent être ajoutées et traitées sur le serveur, qui convertit des données brutes dans un format approprié pour les dashboards de visualisation et l'analyse. Cette infrastructure supporte l'actualisation de la base de données hospitalière, facilitant ainsi l'intégration des données et leur exploitation à travers différents cas d'usage et différentes périodes.

Il est essentiel de noter que les données personnelles des patients ne sont pas utilisées directement ; chaque patient est représenté par un pseudo-code. L'objectif de cet entrepôt de données n'est pas de remonter à des patients individuels mais de fournir une perspective large et populationnelle, permettant des analyses à grande échelle qui respectent la confidentialité et la sécurité des informations des patients.

## **2.3. Transformation et consolidation des données**

La transformation des données hospitalières est essentielle pour les rendre utilisables dans un contexte clinique. D'abord, le système de codification doit être accompagné de descriptions pour permettre l'interprétation par les utilisateurs. Ensuite, les mises à jour ou l'ajout des codes exigent une attention constante pour maintenir l'intégrité des données dans le temps. Enfin, les admissions hospitalières enregistrées sur deux semestres distincts doivent être soigneusement fusionnées pour éviter les redondances. Il est crucial de souligner, comme précisé précédemment, que cet article documente la première phase de l'EDS, un projet en cours de développement. À ce stade, l'EDS intègre principalement des données provenant de sources structurées, telles que les codes ICD-10, qui sont relativement faciles à gérer. Ces sources structurées ne représentent pas la complexité inhérente à la gestion de données non-structurées, comme le texte libre, qui pose un défi significatif en termes de traitement et d'intégration.

## **2.4. Tableau de bord**

Le tableau de bord est un élément central des analyses, rassemblant une variété d'outils conçus pour simplifier et accélérer l'analyse des données, tout en permettant une visualisation efficace. Développé avec PowerBI de Microsoft, il offre des fonctionnalités avancées de filtrage et une diversité de visualisations. De plus, le tableau de bord comprend des tables d'exportation qui facilitent le suivi et la validation des résultats.

## 2.5. IA Générative

Avant de plonger dans l'application concrète de l'IA générative dans notre projet, il est essentiel de définir ce qu'est un Grand Modèle de Langage (Large Language Model (LLM), pour ses initiales en anglais). Les LLMs sont des types avancés d'intelligence artificielle conçus pour comprendre, générer, et interagir en langage naturel. Entraînés à partir d'une vaste quantité de texte, ils peuvent générer des réponses cohérentes et contextuellement pertinentes à des questions variées, facilitant ainsi des interactions dynamiques et naturelles avec les utilisateurs.

Dans le contexte de notre projet, l'IA générative, incarnée par des LLMs, offre une interface utilisateur intuitive permettant des interactions dynamiques avec les bases de données, ouvrant la voie à une exploration plus flexible et approfondie des données disponibles. Bien que l'IA se connecte à l'entrepôt de données, elle n'en fait pas partie intégrante en tant que telle. Cette distinction est importante pour comprendre son rôle en tant qu'intermédiaire entre l'utilisateur et les données stockées. L'architecture de notre solution s'appuie sur un système "Text-to-SQL", qui traduit les questions textuelles des utilisateurs en requêtes SQL exécutables [Kumar *et al.*, 2022, Wang (2023)] localement sur la base de données de l'hôpital.

Le développement de cette solution reste expérimental, explorant de nouvelles façons d'utiliser l'IA comme interface homme-machine. Cette approche expérimentale est motivée par les bénéfices potentiels qu'une solution mature pourrait offrir. Cette idée est soulignée par Jean-Michel Huet (2022) dans son ouvrage 'La révolution de la donnée', où il évoque : 'L'horizon d'une IA capable de répondre à des questions de business intelligence simplement posées par l'homme pourrait briser le plafond de verre représenté par la complexité souvent éprouvée de l'interface homme / machine'. Cette perspective met en lumière le potentiel transformateur de notre approche.

Le développement de cette solution est guidé par la protection des données. Notre système sécurisé utilise uniquement la structure de base de données publique et les requêtes des utilisateurs, garantissant que les données des patients restent sécurisées dans l'hôpital. Pour valider l'exactitude des interactions IA, nous comparons les réponses du LLM avec celles du dashboard traditionnel, assurant ainsi la fiabilité des résultats obtenus.

## 3. Résultats

L'implémentation d'un EDS au Groupe Santé CHC a marqué un tournant dans la manière d'exploiter des données médicales. À travers une série de questions et de cas d'usage initiés par le personnel médical, ce projet démontre la capacité de l'EDS à mobiliser et à transformer les données hospitalières en insights pertinents, appuyant ainsi la prise de décision clinique et la recherche médicale, l'amélioration des soins aux patients et la gestion stratégique de l'hôpital. Quelques résultats obtenus sont présentés dans cette section et se déclinent en deux approches : la visualisation interactive avec un tableau de bord, et l'analyse générative par intelligence artificielle (IA). Il est à noter que l'outil actuellement donne un accès immédiat à des données chiffrées consolidées, cette mise en lumière permet ensuite aux acteurs de terrain de se questionner progressivement et d'obtenir en retour les réponses chiffrées.

### 3.1. Utilisation du tableau de bord

**Étude sur l'appendicectomie.** Au vu de l'intérêt des pédiatres et des chirurgiens, une 1<sup>ère</sup> analyse relative aux appendicectomies a été réalisée en exploitant la profondeur historique des données puisque, entre 2008 et 2023 <sup>1<sup>er</sup> semestre</sup>, 4607 séjours avec un diagnostic d'appendicite aiguë ont été relevés sur l'ensemble des campus hospitaliers. L'outil permet facilement d'établir des points de comparaison, ici 2022 versus 2012, ainsi qu'illustré sur 3 groupes d'âge dans ces graphiques par genre :

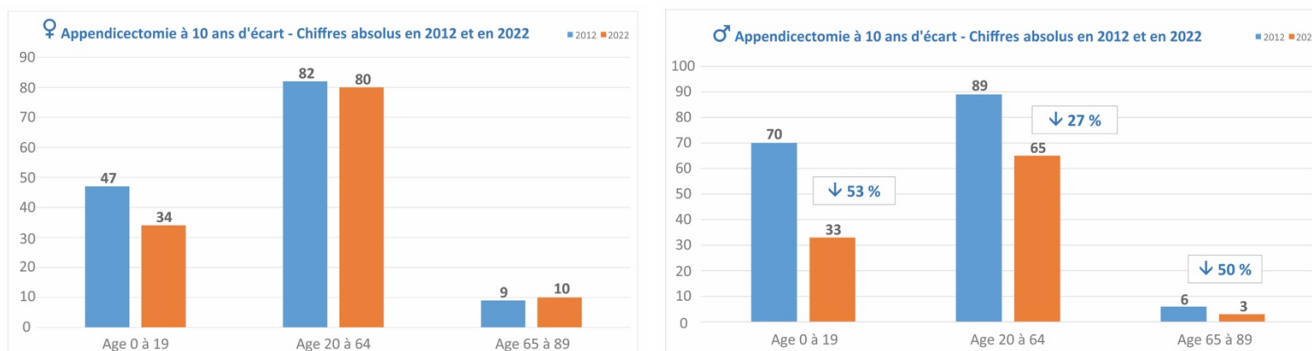


Figure 1. – Comparaison du nombre absolu d'appendicectomie pour 3 groupes d'âge en 2012 et 2022 sur deux sites (gauche et droite).

Il en est de même pour l'évolution de la confirmation anatomo-pathologique :

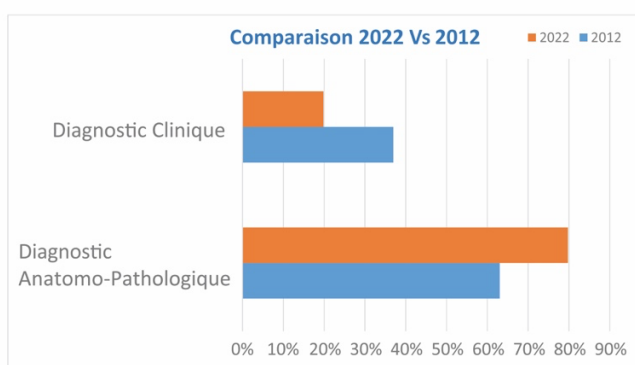


Figure 2. – Comparaison du diagnostic entre 2012 et 2022.

L'Entrepôt permet également de répondre à la volée à des questions complémentaires :

- Réalisation ou non d'une imagerie préopératoire, par CT-Scan et/ou par échographie,
- Pourcentage d'appendicectomie élargie au bas-fond caecal, cette dernière question ayant été posée tout récemment par l'INAMI aux hôpitaux belges.

La granulosité des données permet de remonter les chiffres par campus et par prestataire.

L'évolution prochaine de l'EDS devrait permettre de compléter l'analyse, par exemple en identifiant les patients suspects d'appendicite et qui ont été traités médicalement par antibiothérapie ambulatoire.

**Étude sur la ménisectomie.** Face à une augmentation des ménisectomies, surtout chez les patients de plus de 50 ans, l'INAMI remet en question l'efficacité de ces opérations pour les lésions dégénératives, soulignant que 60% des opérations concernent cette tranche d'âge [Moniteur Belge du 17.04.2023], [Di Prima, C. 2023]. En réponse, l'INAMI a introduit un "seuil de responsabilité", limitant à 45% le taux d'opérations sur les plus de 50 ans. Cette mesure vise à encourager une évaluation plus critique de l'utilité de l'arthroscopie dans le traitement des lésions dégénératives du genou, et à promouvoir des alternatives plus appropriées pour cette catégorie de patients.

Il a été possible de calculer très rapidement l'indicateur et de le contextualiser historiquement par campus et par prestataire. L'ambition ici également est de calculer dans quelle mesure les critères d'*Evidence Based Medicine* publiés par les sociétés savantes sont respectés : imagerie et physiothérapie préalables par exemple.

A gauche, l'évolution du taux, tandis qu'à droite une pyramide des âges cerne rapidement la population en excès, la liste des séjours concernés étant affichée de manière dynamique :

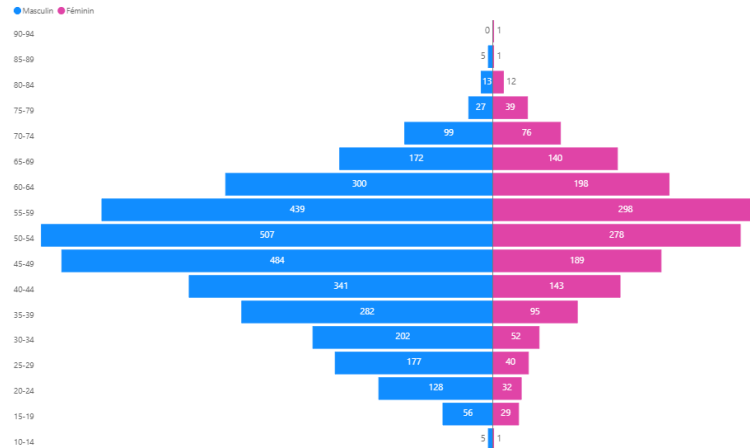
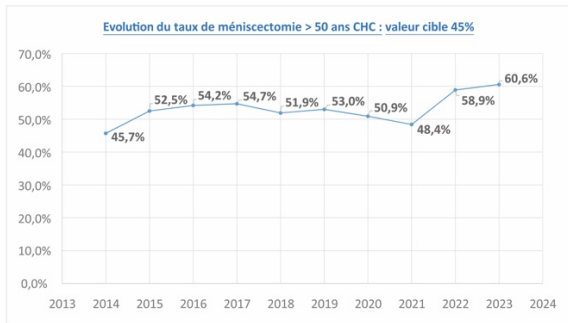


Figure 3 – (A gauche) Évolution du taux de méniscectomie pour les plus de 50 ans. (A droite) Pyramide des âges en fonction du nombre de séjours.

**Étude sur le Diabète.** Pathologie chronique préoccupante en termes de santé publique, le diabète connaît une évolution galopante [Verhaeghe et Van Garderen, 2023], particulièrement en Province de Liège, au point que l'on peut parler d'épidémie. Or, on s'accorde à dire que les chiffres du diabète à l'échelle d'une population sont difficiles à cerner. Nous avons approché cette thématique en exploitant plusieurs sources de données et en explorant l'approche populationnelle, c'est-à-dire en comparant notre population hospitalière à la population nationale.

La consommation d'insuline et d'antidiabétiques oraux (ATC A10 – médicaments antidiabétiques) a ainsi été mesurée et croisée avec les patients qui bénéficient d'une convention spécifique avec l'INAMI pour gérer leur trajet diabétique.

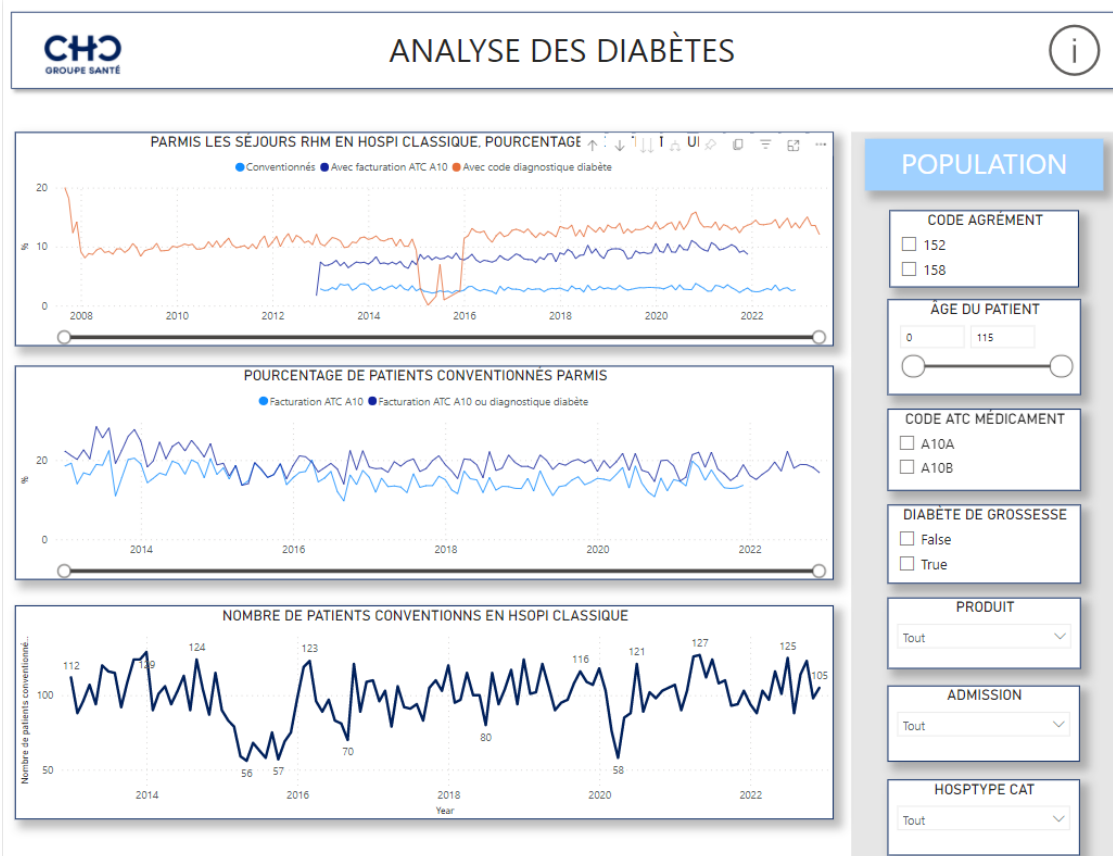


Figure 4. - Tableau de bord pour l'analyse du diabète utilisant le code ATC A10 pour les médicaments antidiabétiques.

**Démographie.** Enfin, nous illustrons brièvement la nécessité bien connue de valider les données et leur exhaustivité au travers d'un accident de données révélé par le volet « Démographie » de l'EDS, qui documente l'ensemble de la population au travers d'indicateurs démographiques tels que l'âge moyen et médian, par genre et par pathologie, sur l'ensemble des années de la période sous revue. Le graphique ci-dessous illustre à la fois l'impact du COVID sur l'âge moyen en 2020, et un pic similaire préalable, expliqué après analyse et recherche par un fichier de données manquant et relatif aux urgences pédiatriques, ce qui provoque artificiellement en 2014 un vieillissement de la population de même importance que le COVID. La partie gauche de la figure présente les données après la correction de l'erreur, illustrant les résultats ajustés.

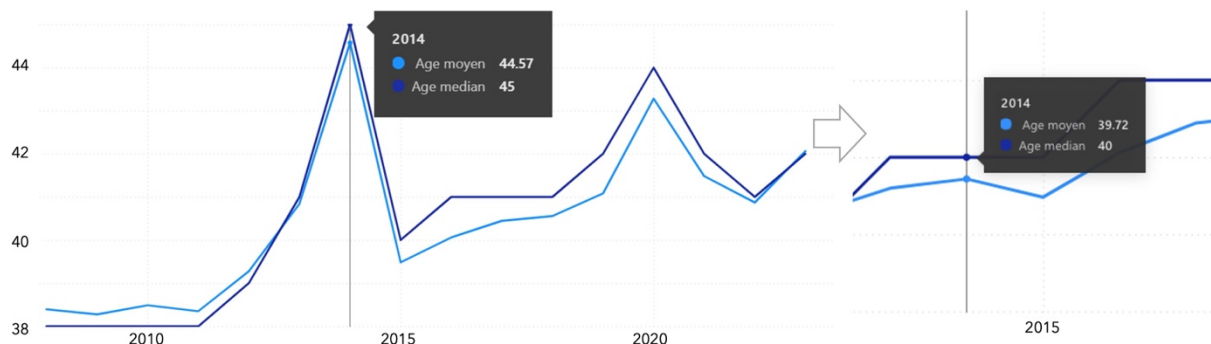


Figure 5. – Évolution de la moyenne/médiane d'âge - Comparaison avant (gauche) et après (droite) correction des données.

### 3.2. Utilisation d'intelligence artificielle générative

**Analyses sur la Post-Mortalité Suite à une Fracture de la Hanche.** En utilisant l'approche de l'IA générative, notre agent médical a pu quantifier et visualiser les tendances de la post-mortalité suite à une fracture de la hanche. L'emploi d'instructions en langage naturel adressées à un LLM a conduit à la génération automatique de graphiques et de statistiques. En outre, les requêtes formulées via le chatbot peuvent être facilement récupérées pour une validation ultérieure, assurant ainsi la fiabilité et la transparence du processus. Notez également dans la Fig. 6, qu'il est possible d'itérer sur les résultats en ajoutant des instructions après avoir reçu une première réponse du LLM, permettant une exploration dynamique des données. Ces résultats ont été validés en faisant une analyse comparative avec ce qui était obtenu via le dashboard.



Dans l'année écoulée, combien de patients ont été admis avec un diagnostique de fracture de hanche ?



Depuis le 12 octobre 2022, il y a eu 557 admissions pour fracture de la hanches, représentant 485 patients. Les codes ICD-10 retenus pour le diagnostique 'fracture de la hanche' sont les codes 'M80' et 'S72'.



Enlève les patients de moins de 65 ans et ceux dont le diagnostique 'fracture de la hanche' n'est pas principal. Donne moi le pourcentage de décès dans cette population selon la durée de séjour



Dans ce cas, il y a 514 séjours pour pour 453 patients. Voici le graphe du taux de décès (%) par durée de séjour:

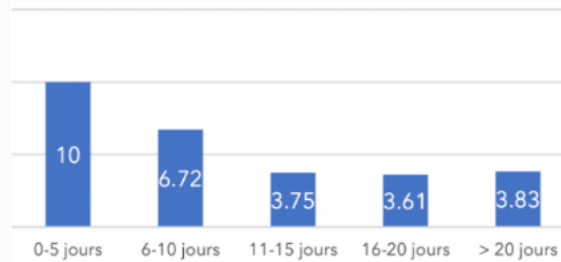


Figure 6. - Taux de mortalité par durée de séjour pour fractures de la hanche. L'interaction se fait en langage naturel.

## Conclusion

Le Groupe Santé CHC a franchi un pas en travaillant sur ce projet de santé numérique. L'adoption d'un EDS et la mise en place d'un tableau de bord efficace ont transformé l'exploitation des données hospitalières, rendant leur utilisation à la fois rapide et intuitive pour répondre aux questions cliniques et administratives. L'utilisation d'un LLM a facilité l'accès aux analyses de données en permettant aux utilisateurs de formuler des questions écrites. Néanmoins, il convient de souligner que cette approche est encore à un stade expérimental et requiert des recherches approfondies pour parvenir à une stabilité permettant une mise en œuvre à grande échelle adéquate. La réussite de ce projet repose sur une collaboration étroite entre data scientists et professionnels de santé, qui ont fourni les cas d'usage pratiques pour orienter le développement de l'EDS et des interfaces utilisateurs pour répondre aux besoins réels du terrain.

Enfin, le caractère réseau de l'EDS a facilité les comparaisons entre les différents hôpitaux du groupe, permettant d'identifier rapidement les différences et de déployer des mesures ciblées. La capacité à comparer les performances et à partager les bonnes pratiques entre cliniques au sein d'un réseau hospitalier souligne la valeur ajoutée de l'entrepôt.

L'enrichissement à court terme de l'EDS par d'autres sources de données : données structurées du Dossier Patient Informatisé, résultats de Biologie Clinique, etc... démultipliera les capacités et la granularité de l'outil.

De plus, le croisement au niveau de la région wallonne ajoutera notamment les données de la 1<sup>ère</sup> ligne de soins, ouvrant ainsi de nouvelles perspectives pour le pilotage du système de santé.



## Glossaire

DRG	Diagnosis Related Groups : Il s'agit d'un système de classification utilisé dans le domaine hospitalier pour catégoriser les cas des patients en différents groupes
INAMI	Institut National d'Assurance Maladie-Invalidité
PAQS	Organisme en Belgique qui vise à promouvoir la qualité et la sécurité dans les soins de santé.
RHM	Résumé Hospitalier Minimum
SPF	Service Public Federal.

## Références

- Di Prima, C. (2023, mai 2). Une décision inédite pour décourager les opérations inutiles du genou chez les plus âgés. *l'Echo*. <https://www.lecho.be/economie-politique//10465000.html>
- Griffier, R., Jouhet, V., Thiessard, F., & Cossin, S. (2020). Identification des verrous et des leviers à la réutilisation secondaire des données dans un établissement de santé. *Revue D Epidemiologie Et De Sante Publique*, 68.
- Hardy, L.R., & Bourne, P.E. (2017). *Data Science: Transformation of Research and Scholarship*.
- Haute Autorité de Santé. (2020). Haute Autorité de Santé. Cartographie des impacts organisationnels pour l'évaluation des technologies de santé. Guide méthodologique. HAS, Saint-Denis La Plaine. [https://www.has-sante.fr/upload/docs/application/pdf/2020-12/guide\\_methodologique\\_impacts\\_organisationnels.pdf](https://www.has-sante.fr/upload/docs/application/pdf/2020-12/guide_methodologique_impacts_organisationnels.pdf)
- Haute Autorité de Santé. (2022). Entrepôts de données de santé hospitaliers en France. HAS, Saint-Denis La Plaine.
- Huet, J. M., & Dugas, F. (2022). *Révolution de la donnée: Les data, la ressource du XXIe siècle*. Pearson.
- Kumar, A., Nagarkar, P., Nalhe, P., & Vijayakumar, S. (2022). *Deep Learning Driven Natural Languages Text to SQL Query Conversion: A Survey* (arXiv:2208.04415). arXiv. <https://doi.org/10.48550/arXiv.2208.04415>
- Lancry, P., O'Connor, R.D., Stempel, D.A., & Raz, M. (2012). Using Health Outcomes Data to Inform Decision-Making. *PharmacoEconomics*, 19, 39-47.
- Moniteur Belge. (2023, avril 17). *Indicateur pour la ménisectomie arthroscopique chez des patients âgés de plus de 50 ans*. <https://www.ejustice.just.fgov.be/>
- Résumé Hospitalier Minimum (RHM)*. (2016, février 18). SPF Santé publique. <https://www.health.belgium.be/fr/sante/organisation-des-soins-de-sante/hopitaux/systemes-denregistrement/rhm>
- Verhaeghe, O., & Van Garderen, F. (2023, juin 24). Un nombre record de Belges atteints de diabète. *l'Echo*. <https://www.lecho.be/economie-politique//10476698.html>
- Wang, P., Shi, T., & Reddy, C. K. (2020, April). Text-to-SQL generation for question answering on electronic medical records. In *Proceedings of The Web Conference 2020* (pp. 350-361).