
MESURER
& AMÉLIORER LA QUALITÉ

RAPPORT

Étude de l'intérêt et de la faisabilité de la mesure d'indicateurs qualité et de sécurité des soins à partir d'entrepôts de données de santé - Annexe 3

Modèle de protocole d'étude

Validé par le Collège le 18 décembre 2024

Descriptif de la publication

Titre	Étude de l'intérêt et de la faisabilité de la mesure d'indicateurs qualité et de sécurité des soins à partir d'entrepôts de données de santé - Annexe 3 Modèle de protocole d'étude
Objectif(s)	Annexe au rapport général « Étude de l'intérêt et de la faisabilité de la mesure d'indicateurs de qualité et de sécurité des soins à partir d'entrepôts de données de santé hospitaliers »
Promoteur(s)	Haute Autorité de santé (HAS)
Auteurs	Agnès Solomiac, Matthieu Doutreligne, Sandrine Morin, Pierre-Alain Jachiet
Validation	Version du 18 décembre 2024

Ce document ainsi que sa référence bibliographique sont téléchargeables sur www.has-sante.fr 

Haute Autorité de santé – Service communication et information
5 avenue du Stade de France – 93218 SAINT-DENIS LA PLAINE CEDEX. Tél. : +33 (0)1 55 93 70 00
© Haute Autorité de santé – décembre 2024

Sommaire

1. Présentation de l'équipe projet	5
1.1. Responsable de traitement	5
1.2. Responsable de la mise en œuvre	5
1.3. Délégué à la Protection des Données	5
1.4. Comité scientifique	5
2. Objectifs et finalités	6
2.1. Contexte, objectif(s) et justification de l'étude	6
2.1.1. Contexte	6
2.1.2. Justification et finalité de l'étude	7
2.1.3. Objectif de l'étude	7
2.2. Justification du respect de l'éthique	7
2.3. Justification de l'intérêt public	7
2.3.1. Finalité et bénéfice du projet	7
2.3.2. Transparence et publication des résultats	7
3. Publication des résultats et valorisation	8
3.1. Livrables	8
4. Méthodologie	9
4.1. Design de l'étude	9
4.1.1. Construire des algorithmes :	9
4.1.2. Mesurer l'indicateur qualité :	9
4.1.3. Valider la mesure :	9
4.2. Description et justification de la population d'étude	9
4.3. Taille de la population	10
4.4. Sources de données	10
4.5. Variables	11
4.6. Préparation de données	11
4.7. Méthodes, traitements et analyses des données	11
4.7.1. Extraction depuis les textes	11
4.7.2. Analyses statistiques	13
4.8. Limites de l'étude	13
4.9. Calendrier prévisionnel et faisabilité du projet	14
4.9.1. Calendrier prévisionnel	14
4.9.2. Faisabilité du projet	14
5. Protection de la vie privée, sécurité et confidentialité des données	15
5.1. Respect des droits des personnes concernées	15

5.2. Confidentialité et sécurité des données	15
5.3. Gestion du risque de réidentification	15
5.4. Support de données	15
Table des annexes	16

1. Présentation de l'équipe projet

1.1. Responsable de traitement

NOM_ES est responsable de traitement.

1.2. Responsable de la mise en œuvre

Ce projet repose sur une collaboration entre **NOM_ES** et la Haute Autorité de Santé. Un contrat de collaboration sera signé entre les deux parties.

NOM_ES

Promoteur du projet

Haute Autorité de santé

La Haute Autorité de santé (HAS) est une autorité publique indépendante à caractère scientifique. Elle vise à développer la qualité dans le champ sanitaire, social et médico-social, au bénéfice des personnes. L'une des missions de la HAS est de mesurer et d'améliorer la qualité dans les hôpitaux, cliniques, en médecine de ville, et dans les structures sociales et médico-sociales.

Pour ce projet, la HAS est

- Co-responsable de la mise en œuvre de l'étude ;
- Le financeur de l'étude. La HAS est financée par le ministère des solidarités et de la santé et l'Assurance Maladie.

Représentant de la HAS

- Haute Autorité de santé : Pr Lionel Collet, Président de la HAS

Membres de l'équipe projet

L'équipe projet est composée de membres de la Mission Data et du service Evaluation et Outils pour la Qualité et la Sécurité des Soins (EVOQSS) de la HAS :

- Matthieu DOUTRELIGNE, chef de projet.
 - Son rôle est de mettre en place, de suivre et de coordonner le projet.
- Agnès SOLOMIAC, biostatisticienne et chargée de projet
 - Son rôle est de mettre en place, de suivre et de coordonner le projet.
- Sandrine MORIN, adjointe à la cheffe du Service EVOQSS.
 - Son rôle est d'apporter son expertise sur la mesure et le développement des IQSS
- Pierre-Alain JACHIET, chef de la Mission Data.
 - Il est le co-responsable hiérarchique de l'équipe projet.
- Laetitia MAY-MICHELANGELI, cheffe du service EVOQSS,
 - Elle est la co-responsable hiérarchique de l'équipe projet

1.3. Délégué à la Protection des Données

Délégué à la Protection des Données de **l'ES**.

1.4. Comité scientifique

Comité scientifique de **l'ES**

2. Objectifs et finalités

2.1. Contexte, objectif(s) et justification de l'étude

2.1.1. Contexte

Le projet « Etude de la faisabilité du calcul automatique d'indicateurs qualité à partir d'Entrepôts de données de santé hospitaliers (EDSH) » est mené par les EDSH et la HAS.

Les indicateurs de qualité et sécurité des soins (IQSS)¹ sont des outils qui permettent d'évaluer un état de santé, une pratique, une organisation ou la survenue d'un événement, ainsi que son évolution dans le temps. Ils peuvent être de différents types : de structure, de processus, ou de résultats. Ils peuvent être mesurés à partir : des dossiers des patients, des bases de données médico-administratives (Programme de médicalisation des systèmes d'information (PMSI), Système National des Données de Santé (SNDS)...), et de questionnaires établissements ou de questionnaires patients.

La Haute Autorité de Santé (HAS) développe et valide des IQSS en établissements de santé (ES) en lien avec les professionnels de santé, les patients et usagers. Ils sont mis à disposition des professionnels de santé pour l'amélioration de la qualité et de la sécurité des soins en établissements de santé.

Aujourd'hui, une grande majorité des IQSS mesurés par la HAS le sont à partir du dossier patient. Cette mesure consiste en un audit rétrospectif portant sur un échantillon aléatoire de séjours réalisés par l'ES. Selon le type des IQSS, jusqu'à 70 séjours par an sont à analyser par ES. Le nombre de dossiers évalués est identique quelles que soient la taille et la catégorie des ES participant au recueil national. En effet, ce nombre de dossiers, relativement faible est un compromis entre une charge de travail acceptable, tous types de structure confondus, et une précision statistique suffisante de l'indicateur estimé à partir de l'échantillon.

Dans ce contexte de ressources professionnels contraintes, la HAS cherche à automatiser le recueil afin de laisser le temps aux professionnels pour se recentrer sur le soin mais aussi leur permettre d'analyser les résultats et mettre en place des plans d'actions. La HAS axe ses travaux sur le développement d'indicateurs mesurés sur l'exhaustivité des soins, ne nécessitant pas de charge de travail supplémentaire pour les ES et pouvant être restitués aux professionnels concernés fréquemment.

C'est dans ce cadre que la HAS souhaite explorer la source de données émergentes qu'est l'Entrepôts de Données de Santé Hospitaliers (EDSH) afin d'accompagner plus loin les ES dans la mise en œuvre de leur politique qualité en leur mettant à disposition des outils et méthodes facilitant le recueil automatique d'indicateurs.

Les EDSH deviennent progressivement une infrastructure centrale importante pour l'appropriation et la compréhension des données cliniques hospitalières par les établissements¹. En décloisonnant les systèmes d'Information hospitaliers au sein d'une base centrale avec un format homogène, et grâce à la mise à disposition d'environnements d'analyses statistiques modernes (python, R), ils permettent aux établissements de santé (majoritairement universitaires) d'apprivoiser la complexité et la richesse de leurs données cliniques. Cette nouvelle infrastructure sert majoritairement pour la recherche. Néanmoins, les établissements considèrent que cette nouvelle source de données permettra d'autres types de réutilisation, notamment dans le cadre du pilotage des soins et de la vigilance. Les équipes cliniques

¹ [rapport_entrepots_donnees_sante_hospitaliers.pdf \(has-sante.fr\)](https://www.has-sante.fr/fr/rapport_entrepots_donnees_sante_hospitaliers.pdf)

sont ainsi intéressées à suivre les indicateurs de qualité et de sécurité des soins de la HAS en dehors des mesures organisées par cette dernière pour disposer d'une mesure adaptée à leur besoin et sur l'exhaustivité de leur cas.

2.1.2. Justification et finalité de l'étude

La HAS cherche à optimiser les mesures de la qualité pilotées par elle. Les EDSH constituant une opportunité intéressante à tester pour évaluer la faisabilité de calcul automatique d'indicateurs à partir de données cliniques, il est important que la HAS puisse jauger de la performance de cette source de données pour l'évaluation de la qualité des soins hospitalier et de son amélioration.

Si l'intérêt d'utiliser les entrepôts est intellectuellement admis, l'équipe clinique de l'ES et la HAS cherchent à évaluer la faisabilité du calcul automatique d'indicateurs de qualité et de sécurité des soins sur les données de l'entrepôt tel qu'elles sont actuellement disponibles. Ceci permettra ensuite de tirer des enseignements sur la structuration nécessaire des données, voire sur la construction des IQSS pour qu'ils soient calculables à partir d'un entrepôt de données cliniques.

2.1.3. Objectif de l'étude

L'objectif principal est l'étude de la performance des EDSH pour calculer automatiquement des indicateurs qualité. Cette performance sera évaluée par la comparaison de la valeur de l'indicateur obtenue sur les données de l'EDSH à aux valeurs déjà calculées par la HAS lors des mesures nationales des IQSS.

Les objectifs secondaires sont de s'assurer des critères de validation de l'indicateur inhérent à la méthodologie de calcul – à savoir au recueil automatique sur l'EDSH et notamment de la stabilité temporelle : capacité de l'outil à produire des résultats cohérents dans le temps.

2.2. Justification du respect de l'éthique

Le projet ne pose pas de problème éthique, pas de stigmatisation d'un groupe spécifique et ne va pas à l'encontre de la morale.

2.3. Justification de l'intérêt public

2.3.1. Finalité et bénéfice du projet

Le projet doit permettre de :

- Disposer de mesures de la qualité plus fréquemment, de façon exhaustive pour un ES donné et sans charge de travail supplémentaire pour les soignants ;
- Accompagner plus loin les établissements dans la mise en œuvre de leur politique qualité en leur proposant de calculer automatiquement des indicateurs qualité ;
- Mettre à disposition les outils et méthodes facilitant la récolte automatique des indicateurs.

A plus long terme, les EDSH et les outils développés dans ce projet pourraient être mobilisés pour faciliter la validation d'IQSS actuellement mesurés sur les données du PMSI.

2.3.2. Transparence et publication des résultats

La HAS prévoit de produire un rapport d'étude, qui détaillera l'ensemble des étapes permettant de répondre à l'objectif principal de l'étude. Ce rapport, ne contenant aucune donnée identifiante, sera disponible au public via une diffusion sur le site internet de la HAS.

3. Publication des résultats et valorisation

Les résultats du projet seront communiqués par la diffusion d'un rapport d'étude complet, publié sur le site internet de la HAS.

Des publications scientifiques sont envisagées par les équipes des EDSH. Des abstracts seront possiblement déposés pour des présentations orales lors de congrès scientifiques.

3.1. Livrables

Le présent travail éditera un guide méthodologique à destination des établissements désireux de mettre en œuvre et tester le calcul automatique d'indicateurs qualité, potentiellement différents de ceux étudiés pour ce travail.

Un outil respectant ces fonctionnalités et déployable aisément au sein d'un environnement d'analyse standard au sein d'un EDSH (de type datalab Jupyter) sera mis à disposition en open source.

4. Méthodologie

4.1. Design de l'étude

Pour répondre à l'objectif principal, une étude descriptive rétrospective sera réalisée pour :

- construire un algorithme PMSI pour identifier la population ciblée par l'indicateur ainsi que des algorithmes spécifiques pour rechercher les deux dates nécessaires au calcul du délai
- et mesurer l'indicateur qualité sur les données de l'EDSH.

4.1.1. Construire des algorithmes :

L'algorithme PMSI permet de sélectionner les séjours/patients de la population évalués par l'indicateur à partir de critères d'inclusion et d'exclusion détaillés dans la fiche descriptive de l'indicateur qualité. Dans un premier temps, cet algorithme sera appliqué aux données du PMSI de l'EDSH.

Dans un second temps, l'ensemble de l'équipe projet, à savoir les cliniciens et data scientists de l'EDSH et l'équipe projet HAS, développera un algorithme permettant de requêter les informations nécessaires à la mesure du numérateur de l'indicateur qualité à partir des données de l'EDSH.

Ce travail consistera d'abord à identifier les bases de données intégrées à l'EDSH qu'il faudra mobiliser pour rechercher ces informations : gestion administrative des malades, PMSI, textes, biologie... Ensuite, les cliniciens spécialistes de la thématique étudiée ainsi que les data scientists de l'EDSH donneront leur avis sur la qualité de la saisie des informations recherchées et sur leur mise à disposition dans l'EDSH. Il conviendra de s'intéresser à la structuration de l'information recherchée. Dans le cas où il serait nécessaire d'extraire des informations depuis les notes cliniques, les étapes standards en traitement automatique du langage (TAL) seront suivies.

4.1.2. Mesurer l'indicateur qualité :

L'indicateur sera mesuré comme la moyenne des délais observés sur chaque séjour/patient de la population d'étude.

4.1.3. Valider la mesure :

Afin de valider la mesure de l'indicateur qualité grâce à l'algorithme construit, la valeur ainsi obtenue sera comparée à la valeur déjà calculée par la HAS lors des mesures nationales des IQSS.

4.2. Description et justification de la population d'étude

La population d'étude correspond aux séjours/patients évalués par l'indicateur.

Critères d'inclusion

Les séjours :

- De patients de 18 ans et plus,
- En hospitalisation complète hors séances,
- D'au moins un jour (1 nuitée),
- Avec en diagnostic principal, un code CIM 10 d'AVC aigu :
 - I60.x (hémorragies sous arachnoïdiennes),
 - I61.x (hémorragies intracérébrales),

- I62.x (autres hémorragies intracrâniennes non traumatiques, sous durale, extradurale, sans précision),
- I63.x (infarctus cérébral) à l'exception de I63.6 (thrombose veineuse non pyogène)
- I64 (accident vasculaire cérébral, non précisé comme étant hémorragique ou par infarctus

Critères d'exclusion

Les séjours :

- avec une erreur (GHM 90)
- avec séances (GHM 28)
- de patients décédés au cours du séjour
- dont la date d'entrée est égale à la date de sortie ou les séjours d'une durée de moins de 24 heures sans nuitée
- faisant l'objet d'une prestation inter-établissements réalisée en tant qu'établissement prestataire (séjour B)
- de patients sortis contre avis médical ou par fuite (Z53.2)
- avec un code d'AIT en DP (G45.x) de n'importe quelle unité médicale

4.3. Taille de la population

L'algorithme permet d'identifier l'exhaustivité des séjours/patients de la population d'étude.

En 2022, 110 124 séjours d'AVC correspondants aux critères d'inclusion et d'exclusion de la campagne (106 783) ont été réalisés dans 1 016 MCO. Ces chiffres ont aussi été calculés pour les **ES participants à cette étude** (cf. Tableau 1).

Tableau 1 Nombre de séjours / patients d'AVC en 2022 (données PMSI MCO) EXEMPLE

Finess	Raison Social	Nombre de sé-jours	Nombre de patients uniques
590780193	CHU de Lille	1 458	1 400
330781196	C.H.U. de Bordeaux	1 605	1 550
..			

4.4. Sources de données

Les données de l'EDSH qui seront mobilisés sont :

Source 1 - PMSI : Les données du PMSI permettent d'identifier la population d'étude (dénominateur de l'indicateur).

Source 2 - Textes cliniques : Le calcul manuel des IQSS s'effectue en analysant l'ensemble du dossier patient. De plus, la typologie des documents dans les entrepôts de données n'est pas tout le temps de bonne qualité et dépend de la qualité du dossier patient. Certains documents peuvent donc contenir un libellé ne correspondant pas à leur contenu. En conséquence, tous les types de documents seront mobilisés pour cette étude.

Source 3 - Métadonnées du PACS : La présence et le timestamp de l'IRM pourront être extraites de ces métadonnées à partir du type d'imagerie dans cette table.

Source 4 - Les données de gestion administrative des malades (GAM) : Les données de flux seront mobilisées pour approcher les dates d'entrées et de sorties des patients à l'hôpital ainsi que dans les différents services d'intérêt.

Source 5 – Pancarte infirmière : Si disponible car pouvant contenir des informations précieuses sur la temporalité de l'arrivée des patients dans un service.

Source 6 – Dossiers de spécialité : Si disponible, les données structurées du dossier de spécialité contiennent les informations nécessaires aux indicateurs retenus.

Source 7 – Prescription et administration des médicaments : Données utiles pour décrire la population d'étude.

4.5. Variables

Les variables nécessaires seront les codes diagnostics et procédures, les codes GHM, les textes cliniques, les métadonnées d'imagerie (PACS), toutes les données structurées du dossier de spécialité.

Pour chaque élément de données mentionnés ci-dessus, les variables importantes pour l'analyse seront : la date et le timestamp de validation de la donnée dans le système d'information, l'identifiant de visite associé à l'élément de données, l'identifiant du patient, l'identifiant du service. Si c'est un document, en supplément les variables suivantes seront nécessaires à l'analyse si disponibles : la spécialité du rédacteur ainsi que le timestamp de rédaction (première et dernière modification idéalement) et de validation du document.

4.6. Préparation de données

L'EDSH mettra à disposition des utilisateurs les données nécessaires au projet (cf. 4.5) sur leur plateforme dans le respect de l'autorisation délivrée par la CNIL. L'accès à cette plateforme se fera via une double authentification.

L'ensemble des données traitées lors de ce projet le seront au sein de la plateforme. Seuls des données agrégées pourront être extraites de la plateforme ; notamment la valeur de l'indicateur et des informations complémentaires (sous forme de tableaux agrégés) permettant de décrire la population d'étude.

4.7. Méthodes, traitements et analyses des données

Les analyses statistiques seront effectuées soit par les personnes par les data scientists de l'ES, sur la plateforme d'analyse de l'EDSH avec les logiciels Python et/ou R, en lien avec les personnes de la HAS travaillant sur le sujet.

4.7.1. Extraction depuis les textes

Les indicateurs retenus nécessitent d'extraire les dates de certains événements médicaux (consultation et examens) depuis les notes cliniques. Les étapes standards en traitement automatique du langage (TAL) pour l'extraction d'information seront suivies :

- **Identification du dataset d'intérêt :** La population d'étude sera identifiée grâce aux critères d'inclusion et d'exclusion PMSI tels que définis dans les fiches descriptives des indicateurs. Les données retenues seront restreintes aux textes relatifs aux séjours ainsi identifiés. De plus, une restriction sera effectuée aux séjours dont les notes cliniques sont effectivement remontées dans l'EDSH (à définir à partir des données). Il sera important de caractériser les séjours ainsi écartés dans un souci de représentativité des séjours étudiés. Une ligne du dataset constitué

correspondra donc à un séjour et les textes lui étant associés. A partir de ce jeu de données principal, un jeu de données de développement sera annoté à la main. Selon les moyens à disposition pour le projet. Idéalement deux cents séjours seront annotés afin de se ramener à des tailles de population similaires pour l'évaluation à celles utilisées pour les IQSS (60 dossiers environ pour chacun des trois jeux de données).

- **Rédaction du guide d'annotation :** Pour chaque indicateur retenu, un guide d'annotation sera rédigé en amont de la consultation des documents afin d'identifier les règles permettant d'annoter le document à la main. L'annotation sera effectuée au niveau du séjour en considérant l'ensemble des documents. Ce travail sera inspiré du guide d'annotation du corpus Merlot ² en proposant pour chaque élément à annoter, une définition précise ainsi que des exemples.
- **Campagne d'annotation avec évaluation inter-annotateurs :** Au moins deux annotateurs par établissement partenaire réaliseront l'annotation des dossiers. L'outil d'annotation retenu sera label studio (cf. Annexe **Erreur ! Source du renvoi introuvable.**) sauf si l'établissement dispose d'un outil interne avec lequel il est plus familier. La campagne s'appuiera sur l'expérience du SEvOQSS ainsi que sur le process d'annotation du corpus Merlot. Les données seront séparées en trois sets correspondants à trois phases de l'annotation.
 - Il y aura une phase d'entraînement des annotateurs lors de laquelle les documents seront annotés systématiquement par chaque annotateur indépendamment puis une annotation de consensus sera obtenue après une phase de discussion entre les annotateurs.
 - Une seconde phase de consensus permettra d'évaluer les accords inter-annotateurs à l'aide de la statistique du Cohen's Kappa.
 - Enfin une phase d'annotation indépendante permettra d'annoter le reste des documents sans faire des annotations multiples.

Découpage en trois sous datasets (entraînement, validation et test) : Le jeu de données sera séparé en deux sous ensemble pour le développement de l'algorithme d'extraction d'information (modèles de TAL) et son évaluation : un jeu d'entraînement sur lequel développer l'algorithme, ainsi qu'un jeu de test servant de gold standard pour l'évaluation finale de l'indicateur selon les critères présentés dans la

- Figure 1. Pour les systèmes de règles, le train set sera également séparé en deux afin de développer et valider les règles sans compromettre le test set. Les séparations se feront au niveau des séjours mais tiendront compte des identifiants patients afin qu'un même patient ne se retrouve pas dans le train ET dans le test set.
- **Développement du modèle de classification (algorithme mixte de classification des compte-rendus) :** Plusieurs modèles supervisés seront testés : entre autres TF-IDF + SVD comme baseline. DrBert³, Hugging-face zero shot classification models. Des modèles à bases de règles seront également testés en observant les données du train set et en utilisant les données du set de validation pour évaluer ces règles sans compromettre l'évaluation finale. L'utilisation des bibliothèques open source edsnlp⁴ et medkit⁵ sera privilégiée pour le développement de ces méthodes à base de règles. Le meilleur modèle d'apprentissage sera sélectionné par cross-validation sur le train set puis réentraîné sur l'ensemble du train set avant d'être utilisé pour la phase de test.

² Campillos, L., Deléger, L., Grouin, C., Hamon, T., Ligozat, A. L., & Névéol, A. (2018). A French clinical corpus with comprehensive semantic annotations: development of the Medical Entity and Relation LIMSI annotated Text corpus (MERLOT). *Language Resources and Evaluation*, 52, 571-601.

³ <https://drbert.univ-avignon.fr/>

⁴ <https://aphp.github.io/edsnlp/latest/>

⁵ <https://heka.gitlabpages.inria.fr/medkit/>

- Evaluations des performances de l'algorithme sur le jeu de test - Validation de la mesure de l'indicateur sur l'EDSH : L'évaluation finale de l'indicateur sera faite à l'aide du test set.

La Figure 1 détaille le processus d'entraînement et de validation :

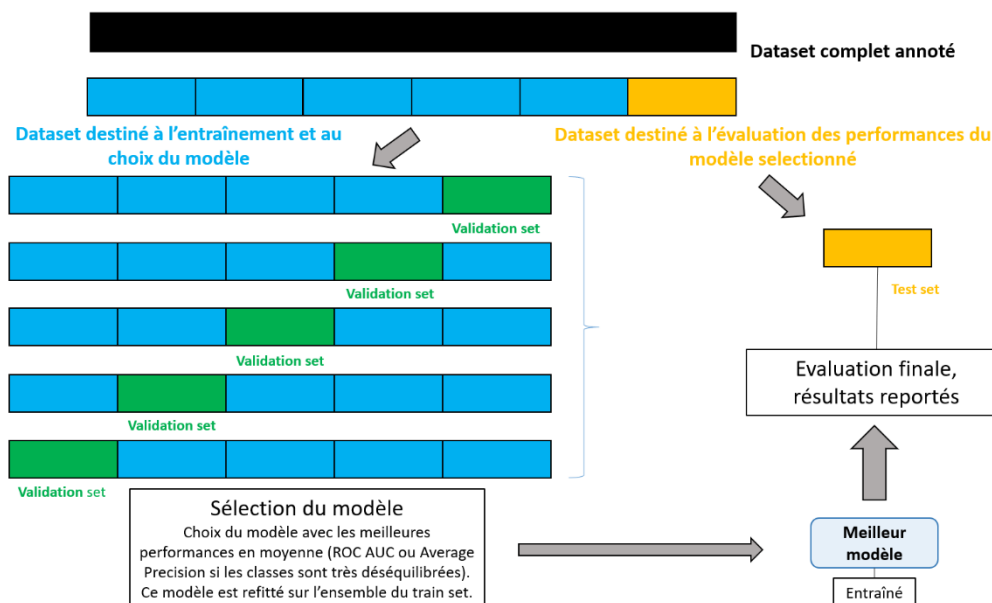


Figure 1 : Procédure de sélection et d'évaluation des modèles de classification

4.7.2. Analyses statistiques

Les variables quantitatives seront décrites en termes de moyenne, écart type, médiane, minimum, maximum, coefficient de variation, quartiles et les faux négatifs seront comparés au reste de la population d'étude à l'aide de tests de variance (ANOVA). Les variables qualitatives ou discrètes seront décrites en termes d'effectif et pourcentage et les faux négatifs seront comparés au reste de la population d'étude à l'aide du test de Chi-2. Le coefficient de kappa sera mesuré pour l'accord inter-annotateurs.

Les valeurs aberrantes et/ou manquantes seront décrites à part.

La population d'étude sera dénombrée et décrite en termes d'âge, de sexe, durée de séjour, type d'AVC (Infarctus cérébral, Hémorragie intracérébrale, Hémorragie sous arachnoïdienne, Hémorragie intracrânienne, AVC de nature indéterminé), mode d'entrée, première unité médicale, mode de sortie.

Les délais de chaque séjour/patient seront décrits comme une variable quantitative. L'indicateur de type délai pourra éventuellement être décliné en termes de taux de délai conforme aux recommandations.

Les distributions des délais moyens estimés par l'algorithme seront estimées de façon non paramétrique par rééchantillonnage (bootstrap) des populations d'études.

4.8. Limites de l'étude

Les limites de l'étude sont surtout liées à la structuration propre d'un entrepôt de données de santé qui complexifie les conclusions universelles et à la méthode :

- La gestion complexe du traitement du langage à partir de textes, potentiellement peu transférable d'un établissement à l'autre.
- La coûteuse vérification par annotation manuelle.

- Dans le cas où la validité de la mesure ne serait pas satisfaisante (par exemple, $\kappa < 0,80$), une réflexion pourra être menée sur l'éventualité de simplifier la cause de conformité. Néanmoins, il faut noter qu'aucun consensus ne semble exister dans la littérature sur le seuil approprié pour un bon κ .

En définitive, toutes les difficultés rencontrées ou interrogations viendront alimenter le rapport final sur la potentielle nécessité d'avoir un format de données standard sur tout un ES pour appliquer les algorithmes construits.

4.9. Calendrier prévisionnel et faisabilité du projet

4.9.1. Calendrier prévisionnel

- Présélection des indicateurs qualité pertinents en interne (cf. Annexe **Erreur ! Source du renvoi introuvable.**) : janvier-février 2023.
- Rédaction du protocole scientifique : février à juin 2023.
- Dépôts des projets de recherche aux comités scientifiques et éthiques des établissements partenaires : juin-juillet 2023.
- Validation des comités et signature des conventions : septembre 2023.
- Campagne d'annotation : novembre-décembre 2023
- Mise en œuvre des études sur entrepôts : décembre-janvier 2024.
- Rédaction du rapport de rapport de résultat, validation par le Collège de la HAS et publication : ~ juin 2024.

4.9.2. Faisabilité du projet

Pour exploiter au mieux les données de l'entrepôt de données de santé, l'équipe multidisciplinaire de l'établissement composée du clinicien co-porteur du projet ainsi que des data scientists capables de manipuler les données collabore avec la HAS.

A la HAS, le service EvOQSS a une grande expérience dans le développement des indicateurs de qualité et de sécurité des soins, compte-tenu de la mission qu'il conduit déjà au niveau national, et la Mission Data a développé une grande expertise sur les EDSH.

5. Protection de la vie privée, sécurité et confidentialité des données

5.1. Respect des droits des personnes concernées

Comme spécifié par le référentiel CNIL sur les EDSH, les patients seront informés de cette réutilisation de leurs données. Cette information se fera par la publication du protocole d'étude sur le site des ES partenaires.

5.2. Confidentialité et sécurité des données

5.3. Gestion du risque de réidentification

5.4. Support de données

Table des annexes

EXEMPLES

Annexe 1.	Listes des IQSS que la HAS collecte au niveau national (exemple)	17
Annexe 2.	Choix de l’outil d’annotation	18
Annexe 3.	Fiches descriptives (exemple)	19

Annexe 1. Listes des IQSS que la HAS collecte au niveau national (exemple)

Thème	Indicateur
Santé mentale – Hospitalisation à temps plein	Repérage et proposition d'aide à l'arrêt des addictions
Santé mentale – Hospitalisation à temps plein	Evaluation gastro-intestinale chez les patients adultes
Santé mentale – Hospitalisation à temps plein	Evaluation cardio-vasculaire et métabolique chez les patients adultes
Santé mentale – Hospitalisation à temps plein	Qualité de la lettre de liaison à la sortie
Santé mentale – Hospitalisation à temps plein	Evaluation et prise en charge de la douleur somatique
Santé mentale - Ambu	Vigilance de l'état cardio-vasculaire et métabolique chez les patients adultes
Santé mentale - Ambu	Vigilance sur le transit intestinal chez les patients adultes
Santé mentale - Ambu	Evaluation et prise en charge des addictions chez les patients adultes
Santé mentale - Ambu	Lien entre l'hôpital et la médecine de ville
...	

Annexe 2. Choix de l'outil d'annotation

Dans le cadre d'un autre projet de traitement du langage nécessitant l'annotation de documents (classification des commentaires E-SATIS), la HAS a comparé différents outils d'annotation. Finalement, l'outil retenu a été [label studio](#), un outil dont la version freemium peut-être déployée aisément à l'aide de docker. Les avantages et inconvénients sont listés ci-dessous :

Avantages

- Permet le paramétrage de l'interface d'annotation (e.g., mixer les niveaux de classification : texte, mots (Named Entity Recognition, Part Of Speech, ...))
- Une fenêtre pour filtrer les textes à annoter
- Export correct en plusieurs formats avec des métadonnées (temps d'annotation notamment)

Inconvénients

Mauvaise gestion des utilisateurs (pas de niveau d'accès, tous peuvent changer les annotations des autres annotateurs)

L'annotateur A peut voir les annotations de l'annotateur B. Cela empêche l'annotation double.

Afin de pallier les inconvénients, nous proposons de déployer différents dockers pour chaque annotateur. Le paramétrage de cette solution est actuellement mis en œuvre à la HAS et nous pouvons le transférer au sein des entrepôts partenaires si ceux-ci ne disposent pas de solution interne.

La grille de comparaison avec d'autres outils d'annotation open source est fournie comme fichier annexe (PROTOCOL_EDSH_IQSS-Annexe_2-grille_outils_annotation.xlsx)

Annexe 3. Fiches descriptives (exemple)

Fiche descriptive de l'information complémentaire : « Délai entre l'arrivée dans l'établissement et la 1ère imagerie cérébrale à visée diagnostique » 2023	
Définition	Cette information complémentaire évalue le taux de dossiers de patients permettant de déterminer le délai entre l'arrivée dans l'établissement et la 1 ère imagerie cérébrale à visée diagnostique, si ce n'est pas celle qui a confirmé le diagnostic d'AVC.
Justification	Les patients, pour lesquels un diagnostic d'accident vasculaire cérébral aigu est évoqué, doivent avoir un accès prioritaire à une imagerie cérébrale (IRM ou scanner). Celle-ci est indispensable pour confirmer ou infirmer un diagnostic d'AVC aigu et pour distinguer un infarctus cérébral, d'une hémorragie ou thrombose cérébrales et prendre les décisions thérapeutiques recommandées (1, 2-8). Des protocoles de prise en charge doivent être formalisés avec le service de radiologie pour un accès 7/7 et 24/24 (1,2). Pour les établissements sans unité neurovasculaire, l'accès à la téléexpertise (téléAVC) doit être opérationnel. Le délai entre l'arrivée dans l'établissement et l'imagerie contribue au délai entre l'arrivée dans l'ES et la thrombolyse des infarctus cérébraux « time door to needle ». Ce délai qui doit être le plus court possible est considéré par toutes sociétés savantes comme un indicateur de qualité des pratiques et organisations nécessaires à la prise en charge des infarctus cérébraux. De même, le pronostic des hémorragies cérébrales est associé à la rapidité de mise en œuvre d'une prise en charge adaptée, permettant de suivre l'état neurologique, les fonctions vitales et de contrôler le processus hémorragique et ses conséquences. Cette notion est mise en exergue dans les recommandations les plus récentes (3-8). 1 HAS 2009. Recommandation pour la prise en charge d'un AVC. https://www.has-sante.fr/jcms/c_830203/fr/accident-vasculaire-cerebralprise-en-charge-precoc-alerte-phase-prehospitaliere-phasehospitaliere-initiale-indications-de-la-thrombolyse 2 DGOS (2012). Circulaire DGOS/R4/PF3/2012/106 du 6 mars 2012 relative à l'organisation des filières régionales de prise en charge des patients victimes d'AVC. https://solidarites-sante.gouv.fr/fichiers/bo/2012/12-04/ste_20120004_0100_0081.pdf 3 European Stroke Organisation, ESO. (2013). Recommendations to establish a stroke unit and stroke Centres. ESO 2013, http://stroke.ahajournals.org/content/44/3/828 4 American Heart Association/American Stroke Association, ASA & AHA. (2015). Guidelines for the management of spontaneous intracerebral haemorrhage from American Heart Association/American Stroke Association.

	Stroke. 2015; 46:2032-2060. DOI: 10.1161/STR.0000000000000069 https://www.ahajournals.org/doi/10.1161/str.0000000000000069 5 National Institute for care excellence, NICE. (2016). Stroke in adults Quality standard [QS2] Published: 29 June 2010 Last updated: 12 April 2016 6 American Heart Association/American Stroke Association, ASA & AHA. (2019). Guidelines for the Early Management of Patients with Acute Ischemic Stroke: 2019 Update to the 2018. A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association https://www.ahajournals.org/doi/10.1161/STR.0000000000000211 HAS/DAQSS/SEvOQSS - Prise en charge initiale d'un accident 13/44 V.05/2023 vasculaire cérébral aigu en MCO (AVC MCO) - Fiches descriptives 2023 7 National Institute for care excellence (NICE). 2019. Stroke and transient ischaemic attack in over 16s: diagnosis and initial management. Guidelines https://www.nice.org.uk/guidance/ng128/resources/strokeand-transient-ischaemic-attack-in-over-16s-diagnosis-and-initialmanagement-pdf-66141665603269
Utilisations	Pilotage interne de la qualité
Type de mesure	Information complémentaire Taux
Numérateur	Nombre de dossiers patients dont les informations permettent de calculer le délai entre l'arrivée dans l'établissement et la première imagerie cérébrale à visée diagnostique, qui n'est pas l'imagerie qui a confirmé le diagnostic • La date et l'heure d'arrivée dans l'établissement Et • La date et heure de la première imagerie cérébrale à visée diagnostique d'un AVC qui n'est pas celle qui a confirmé le diagnostic d'AVC aigu. Le délai est calculé à partir des dossiers correspondant aux critères d'inclusion et d'exclusion ci-dessous.
Dénominateur	Les dossiers répondant aux critères d'inclusion et d'exclusion ci-dessous.
Échantillon	L'indicateur est calculé sur un échantillon aléatoire de 50 dossiers tirés au sort sur l'année 2022, débutés et terminés en 2022
Critères d'inclusion	L'échantillon principal est constitué de séjours de patients adultes • En hospitalisation complète hors séances, • D'au moins un jour (1 nuitée) • Commencés et terminés en 2022. • Avec en diagnostic principal, un code CIM 10 d'AVC aigu4 : - o I60.x (hémorragies sous arachnoïdiennes), - o I61.x (hémorragies intracérébrales),

	- o I62.x (autres hémorragies intracrâniennes non traumatiques, sous durale, extradurale, sans précision), - o I63.x (infarctus cérébral) à l'exception de I63.6 (thrombose veineuse non pyogène) - o I64 (accident vasculaire cérébral, non précisé comme étant hémorragique ou par infarctus) • En provenance de leur lieu de résidence (domicile ou structure d'hébergement médico-social) et tous modes de sortie hors décès.
Critères d'exclusion	• Les séjours avec un erreur (GHM 90) • Les séjours avec séances (GHM 28) • Les séjours de patients âgés de moins de 18 ans • Les séjours de patients décédés au cours du séjour • Les séjours dont la date d'entrée est égale à la date de sortie ou les séjours d'une durée de moins de 24 heures sans nuitée • Les séjours faisant l'objet d'une prestation inter-établissements réalisée en tant qu'établissement prestataire (séjour B) • Les séjours de patients sortis contre avis médical ou par fuite (Z53.2) • Les séjours avec un code d'AIT en DP de n'importe quelle unité médicale (G45.x)
Critères d'exclusion secondaires	• Dossiers pour lesquels le diagnostic radiologique d'AVC aigu est antérieur l'admission • Dossiers pour lesquels le diagnostic radiologique d'AVC aigu est confirmé par la première imagerie cérébrale réalisée dans ou pour votre établissement
Mode d'évaluation de l'information complémentaire	Il convient de trouver dans le dossier des patients arrivant du domicile, la trace des informations permettant de déterminer le délai entre la date et heure d'arrivée dans l'établissement et la première imagerie cérébrale à visée diagnostique d'un AVC aigu, si elle n'est pas celle qui a confirmé le diagnostic. L'évaluation des dossiers est réalisée à l'aide d'un questionnaire. • Une trace de la date et heure de l'arrivée dans l'ES ET • La date et heure de la première imagerie à visée diagnostique d'un AVC, qui n'est pas celle qui a confirmé le diagnostic

Retrouvez tous nos travaux sur
www.has-sante.fr

