Nº 11-522-x au catalogue
ISSN : 1709822-X

Recueil du Symposium de 2021 de Statistique Canada
Adopter la science des données en statistique officielle pour répondre aux besoins émergents de la société

Exploiter le traitement du langage naturel et l'apprentissage automatique pour améliorer la détermination des résultats en matière de santé liés aux opioïdes dans l a national hospital care survey

par Amy M. Brown et Nikki Adams

Date de diffusion : le 22 octobre 2021

Statistique Canada Statistics Canada

Canada

Recueil du Symposium de 2021 de Statistique Canada
Adopter la science des données en statistique officielle pour répondre aux besoins émergents de la société

Exploiter le traitement du langage naturel et l'apprentissage automatique pour améliorer la détermination des résultats en matière de santé liés aux opioïdes dans la National Hospital Care Survey

Amy M. Brown, MSP, et Nikki Adams, Ph. D.[1]

Résumé

Le National Center for Health Statistics a reçu un financement pour deux projets visant à accroître la capacité en données et à rendre compte de la crise de santé publique des opioïdes aux États-Unis. Ceux-ci consistent à mettre au point des algorithmes qui font appel à toutes les données disponibles, structurées ou non, de la National Hospital Care Survey (NHCS) de 2016 pour mieux connaître la dépendance aux opioïdes et la présence de troubles concomitants (coexistence de troubles liés à l'utilisation de substances et de problèmes de santé mentale). Nous décrivons la démarche d'élaboration de ces algorithmes et les leçons tirées de l'intégration de méthodes de science des données comme le traitement du langage naturel dans la production des statistiques officielles. Il sera également question des mesures visant à rendre accessibles aux chercheurs les algorithmes et les fichiers de données analytiques.

Mots clés : opioïdes; troubles concomitants; science des données; traitement du langage naturel; soins hospitaliers

1. Introduction

1.1 Contexte du projet

Le National Center for Health Statistics (NCHS), qui relève des Centers for Disease Control and Prevention (CDC), compte parmi les 13 organismes statistiques fédéraux et est le principal organisme en matière de statistique de la santé aux États-Unis. Les systèmes de collecte de données comprennent les registres de l'état civil, les enquêtes auprès de la population et les enquêtes auprès des soignants. La National Hospital Care Survey (NHCS) est une enquête conçue pour fournir des données sur l'utilisation des services de santé en milieu hospitalier. Dans les exercices de 2018 et de 2019, le NCHS a reçu des fonds du secrétariat du Patient-Centered Outcomes Research Trust Fund (PCORTF) du département de la Santé et des Services sociaux pour deux projets destinés à renforcer les capacités en matière de données et à intégrer les méthodes de science des données pour mieux rendre compte de la crise de santé publique des opioïdes. Le but est d'améliorer le dénombrement et la caractérisation des contacts hospitaliers liés aux opioïdes. Ces projets consistaient à concevoir deux jeux algorithmiques exploitant toutes les données disponibles, structurées ou non, dans l'information de plus en plus abondante et complexe de la NHCS. Le premier de ces algorithmes est là pour recenser les contacts hospitaliers avec consommation passée ou présente d'opioïdes et pour reconnaître le type d'opioïdes consommés. Le second vise à recenser les utilisateurs d'opioïdes présentant des troubles concomitants et, plus particulièrement, des troubles liés à l'utilisation de substances (TUS) avec des problèmes de santé mentale (PSM). Ces algorithmes améliorés sont là pour relever les cas qui pourraient avoir été manqués par les algorithmes du passé où on comptait seulement, dans les données d'enquête présentées, sur les codes médicaux pour reconnaître les problèmes d'intérêt en santé des comportements.

[1]Amy M. Brown, MSP, National Center for Health Statistics, 3311, chemin Toledo, États-Unis, 20782 (wri1@cdc.gov); Nikki Adams, Ph. D., National Center for Health Statistics, Centers for Disease Control and Prevention, 3311, chemin Toledo, États-Unis, 20782 (oxf7@cdc.gov). Tous les points de vue exprimés dans cet article appartiennent aux auteurs et ne reflètent pas nécessairement les vues des Centers for Disease Control and Prevention ni du National Center for Health Statistics.

1.2 Sources de données

L'échantillon actuel de la NHCS comprend 608 hôpitaux non institutionnels et non fédéraux comptant six lits d'hospitalisation avec préposés ou plus. Les hôpitaux participants sont appelés à livrer leurs données sur l'ensemble des hospitalisations et des visites en salle d'urgence pour chaque année civile. Pour soumettre leurs données, les établissements hospitaliers peuvent choisir entre deux grands supports de communication électronique : (1) formulaire de facturation uniforme (UB)-04 (demandes administratives) et (2) dossier de santé électronique (DSE). Les données présentées sont une information largement diversifiée, dont des données structurées comme les codes médicaux normalisés (codes de diagnostic et d'intervention, par exemple) et des données non structurées sous forme de notes cliniques des soignants. Les données des deux projets financés par le PCORTF viennent de la NHCS de 2016 qui rend compte d'environ 9,6 millions d'hospitalisations et de visites en salle d'urgence dans 158 hôpitaux.

2. Méthodes

2.1 Processus de développement de l'algorithme

La démarche de développement de l'algorithme a débuté par la création de définitions de cas pour chacun des deux algorithmes en question. La définition applicable à l'algorithme amélioré d'identification de la dépendance aux opioïdes comportait des critères permettant de reconnaître la consommation passée ou présente de l'ensemble des opioïdes d'ordonnance et illicites et autres substances aux effets assimilables à ceux des opioïdes comme le kratom. Quant à la définition du cas de l'algorithme des troubles concomitants, elle est assortie de critères permettant de constater les troubles de consommation passée ou présente de toutes les substances avec certains problèmes de santé mentale comme l'anxiété, la dépression, le trouble obsessionnel compulsif, les idées suicidaires et les troubles liés aux traumatismes et aux facteurs de stress.

Les algorithmes comprenaient deux éléments puisant dans toutes les données disponibles relativement aux divers contacts pour cerner les cas. Le premier relevait les codes et mots clés médicaux applicables dans tous les domaines avec le codage médical et l'étiquetage ou la description applicable. Avec le second, on annotait ou étiquetait un échantillon de notes cliniques pour alimenter en données la mise au point de techniques de traitement du langage naturel (TLN) et d'apprentissage automatique permettant d'analyser du texte non structuré de notes cliniques. Le volet TLN consistait à décomposer en phrases le texte des notes cliniques et à passer celles-ci en revue pour détecter plusieurs types d'exclusions, qu'il s'agisse d'exclure des opioïdes ou des mentions « TUS » (troubles de consommation) ou « PSM » (problèmes de santé mentale) liées à des dates postérieures à 2016. Il y avait également détection des négations pour voir si une mention de consommation, de diagnostic ou de maladie dans un rapport médical narratif était affirmative ou négative (« absence d'antécédents d'anxiété », par exemple). On relevait enfin les fautes d'orthographe dans les mentions d'opioïdes en appliquant un modèle d'apprentissage automatique appelé reconnaissance d'entités désignées, après un entraînement du modèle en vue de reconnaître les noms de médicaments, puis de comparer les noms relevés aux noms bien orthographiés.

Le cadre méthodologique applicable à l'algorithme amélioré d'identification de la dépendance aux opioïdes est décrit en détail dans un rapport publié (White et coll., 2021). Un rapport semblable à paraître traitera en détail du cadre méthodologique applicable à l'algorithme des troubles concomitants. À l'heure actuelle, nous réalisons une étude de validation du rendement de tels algorithmes améliorés. Les résultats de cet exercice en cours nous permettront d'apporter des modifications et de produire les algorithmes sous leur forme définitive.

2.2 Élément des codes médicaux

L'année d'enquête 2016 est la première où les hôpitaux participant à la NHCS avaient la possibilité de présenter les données non structurées des notes cliniques médicales sauvegardées dans leurs dossiers électroniques de santé; c'était l'occasion d'obtenir de nouveaux types de codes médicaux indisponibles par le passé. L'élément des codes médicaux pour les deux algorithmes visait des codes choisis dans les systèmes de codage normalisé de diagnostic, de médication, d'intervention et d'essai de laboratoire. Dans certains cas, on s'est aussi intéressé à des mots clés choisis dans l'étiquetage ou la description des codes, là où l'hôpital employait un système de codage non normalisé. On peut

consulter les listes finales de codes et de termes médicaux de recherche au site Web du centre de données de recherche du NCHS (NCHS, 2020; NCHS, 2021).

2.3 Élément TLN

L'introduction des données de dossier de santé électronique (DSE) dans la NHCS de 2016 a également permis de disposer des données non structurées des notes cliniques pour la première fois dans l'histoire de cette enquête. Ces notes peuvent nous renseigner plus précisément sur la nature des opioïdes consommés comparativement aux codes médicaux normalisés attribués aux contacts hospitaliers. De même, les recherches dans les notes cliniques sous les rubriques « Antécédents médicaux » et « Antécédents sociaux » peuvent aider à relever les patients ayant reçu un diagnostic clinique de troubles concomitants ou de problèmes de santé mentale avant le contact hospitalier.

La méthode générale du volet TLN reposait principalement sur des règles avec un volet d'apprentissage automatique en langage Python permettant de relever les fautes d'orthographe des appellations d'opioïdes. Au début, nous avons réuni des mots et des locutions clés pour les catégories d'intérêt, après quoi nous avons affiné le tout. Il s'agissait d'extraire le texte contenu dans les notes à la suite d'un premier filtrage des contacts hospitaliers. Il fallait, par exemple, éliminer les notes générales caractéristiques des feuillets d'information des patients. Nous décomposions alors le texte en phrases et procédions à des exclusions au niveau des phrases (s'il s'agissait, par exemple, de phrases sur les antécédents familiaux). Des recherches par mots clés se faisaient ensuite. Pour la détection de la dépendance aux opioïdes seulement, la technique de reconnaissance d'entités désignées servait à relever les noms de médicaments ne figurant pas déjà sur notre liste (fautes d'orthographe). Suivait un mode automatique de comparaison orthographique des termes candidats avec les appellations connues des opioïdes et autres médicaments. Si l'orthographe était ressemblante, des annotateurs humains devaient confirmer par la suite s'il y avait bel et bien faute d'orthographe. En plus de l'application d'un modèle de base en anglais du progiciel Spacy (spacy, https://spacy.io), le modèle de reconnaissance des entités désignées a été entraîné à l'interne à l'aide d'un ensemble de données annotées de la NHCS de 2016. À des fins d'exclusion par occurrence négative (« le patient nie abuser des opioïdes », par exemple), tous les appariements avec les mots clés passaient par un filtre de détection des négations appelé Negex (Chapman et coll., 2001). Les appariements sans négation étaient enfin appliqués aux catégories d'ensemble des variables pour dégager l'ensemble final de données. Pour plus de renseignements, voir White et coll., 2021.

L'intégration de ces notes cliniques à la NHCS de 2016 a toutefois aussi présenté de nouveaux défis. Il n'y a que pour 8,7 % de tous les contacts hospitaliers qu'au moins un enregistrement de notes médicales était disponible à des fins de traitement TLN. Les notes médicales étaient fournies à titre facultatif, car il est plus difficile pour les hôpitaux de les extraire. Les notes présentées étaient entachées de fautes d'orthographe, de formes tronquées, de blancs et de variantes. Leur format était variable aussi (texte libre ou format XML), ce qui exigeait beaucoup de nettoyage et de reformatage. Enfin, l'élément TLN devait voir le jour dans un environnement informatique clos qui respectait les règles applicables de sécurité des données, ce qui venait limiter la puissance disponible de traitement logiciel et de calcul. Ce cadre fermé était par ailleurs difficile d'accès en période de pandémie de COVID-19, car il était inaccessible depuis le domicile en situation de télétravail. Pour tenir compte de cette fermeture et du manque de disponibilité de moyens logiciels, comme solution pour l'annotation le personnel de projet devait s'appuyer davantage sur un logiciel conçu à l'interne.

3. Résultats

3.1 Résultats de l'annotation

L'ensemble de données d'annotation a été divisé en ensembles de développement TLN et en ensembles d'évaluation de rendement du traitement avec un ensemble de chaque type pour l'algorithme de la dépendance aux opioïdes, de l'algorithme des problèmes de santé mentale (PSM) et de l'algorithme des troubles de consommation (TUS). Les résultats de l'annotation traduisent le rendement des éléments codes médicaux et TLN et de l'algorithme d'ensemble. Nous avons combiné les deux éléments pour établir les distinctions nécessaires (présence ou absence d'opioïdes, TUS et PSM) selon les indications des annotateurs. Tous les contacts annotés comportaient des notes cliniques et les deux

éléments de l'algorithme d'ensemble étaient également applicables à tous les contacts dans l'ensemble de données correspondant. Les tableaux 3.1-1 à 3.1-3 présentent les valeurs de mesure des résultats suivantes :

- rappel : pourcentage de positifs dûment relevés parmi tous les vrais positifs, ce qu'on appelle la sensibilité;
- précision : pourcentage de positifs relevés qui sont de vrais positifs, ce qu'on appelle la valeur prédictive positive (VPP);
- F1 : moyenne harmonique du rappel et de la précision constituant une mesure commune du rendement de l'algorithme;
- CCM : coefficient de corrélation de Matthew aussi appelé coefficient phi de Pearson, une mesure d'équilibre sur les vrais et les faux négatifs et positifs.

Dans le cas de l'algorithme amélioré d'identification de la dépendance aux opioïdes, les notes rappel, F1 et CCM étaient basses pour l'élément des codes médicaux, alors que l'algorithme d'ensemble (codes médicaux et TLN) présentait le meilleur résultat dans l'ensemble avec des notes F1 et CCM de 92,5 % et 0,77 (tableau 3.1-1). Ce résultat était à prévoir, parce que les notes cliniques font souvent état d'opioïdes, et plus particulièrement de produits thérapeutiques, pour lesquels il n'y a aucun code médical correspondant dans les zones de codage. Dans le cas de l'algorithme des troubles concomitants, les deux éléments combinés réussissaient le mieux à relever les problèmes de santé mentale (PSM) avec un F1 de 95,2 % et un CCM de 0,82 (tableau 3.1-2). Pour les troubles de consommation (TUS), l'élément des codes médicaux était le meilleur (F1 de 94,9 % et CCM de 0,80), mais l'algorithme d'ensemble était toujours d'un bon rendement avec un F1 (90,2 %) et un CCM (0,80) relativement élevés (voir le tableau 3.1-3).

Tableau 3.1-1
Rendement de l'algorithme amélioré d'identification de la dépendance aux opioïdes par rapport à l'ensemble de données annotées et selon les éléments

	Élément des codes médicaux	Élément TLN	Algorithme d'ensemble (codes et TLN)
Rappel	25,5 %	94,8 %	96,9 %
Précision	96,9 %	88,5 %	88,6 %
F1	20,4 %	91,5 %	92,5 %
CCM	0,30	0,74	0,77

NOTES : TLN, traitement du langage naturel; F1, moyenne harmonique rappel-précision; CCM, coefficient de corrélation de Matthew.
SOURCE : National Center for Health Statistics, National Hospital Care Survey, 2016.

Tableau 3.1-2
Rendement de l'algorithme des troubles concomitants en détection des PSM par rapport à l'ensemble de données annotées et selon les éléments

	Élément des codes médicaux	Élément TLN	Algorithme d'ensemble (codes et TLN)
Rappel	86,7 %	74,7 %	93,3 %
Précision	99,2 %	96,6 %	97,2 %
F1	92,5 %	84,2 %	95,2 %
CCM	0,77	0,58	0,82

NOTES : TLN, traitement du langage naturel; F1, moyenne harmonique rappel-précision; CCM, coefficient de corrélation de Matthew.
SOURCE : National Center for Health Statistics, National Hospital Care Survey, 2016.

Tableau 3.1-3
Rendement de l'algorithme des troubles concomitants pour la détection des TUS par rapport à l'ensemble de données annotées et selon les éléments

	Élément des codes médicaux	Élément TLN	Algorithme d'ensemble (codes et TLN)
Rappel	91,6 %	90,2 %	99,3 %
Précision	98,5 %	81,1 %	82,6 %
F1	94,9 %	85,4 %	90,2 %
CCM	0,90	0,70	0,80

NOTES : TUS, troubles liés à l'utilisation de substances; TLN, traitement du langage naturel; F1, moyenne harmonique rappel-précision; MCC, coefficient de corrélation de Matthew.
SOURCE : National Center for Health Statistics, National Hospital Care Survey, 2016

L'élément des codes était des plus précis, mais le rappel était généralement moindre. Signalons en particulier que, comme la définition de cas était la consommation de tout opioïde dans le cas de la dépendance aux opiacés, la recherche par mots clés donnait de bons résultats dans ce cas, alors que fréquemment il n'y avait pas de code de facturation correspondant à cette consommation. On obtenait de faux négatifs et de faux positifs pour l'élément TLN à cause d'une diversité de facteurs. Nous tâchons toujours de mieux comprendre ces raisons. Nous poursuivons l'analyse des erreurs et continuons à analyser les résultats d'une étude d'évaluation en suivi. Il reste qu'un premier examen dégage des raisons pour les faux négatifs, comme les interactions médicamenteuses caractérisées comme « abus » par le code de diagnostic, mais comme « consommation » dans les notes, là où cette consommation, selon notre définition de cas, n'atteignait pas le niveau des troubles liés à l'utilisation de substances. Une source fréquente de faux positifs dans l'élément TLN était la nicotine, car la définition de cas pour la consommation TUS de nicotine n'était pas la même entre le stade de l'annotation et la période d'exécution de l'algorithme. Un ancien fumeur était exclu d'abord et inclus ensuite. Comme l'état d'« ancien fumeur » était mentionné dans les notes mais sans recevoir de code de diagnostic pour les antécédents personnels de dépendance à la nicotine, l'élément TLN relevait les anciens fumeurs plus souvent que l'élément des codes médicaux. L'analyse des erreurs et des résultats de notre étude de validation se poursuit.

3.2 Application d'algorithmes à la NHCS de 2016

Après annotation, les algorithmes améliorés ont été appliqués à la base de données de la NHCS de 2016. Les tableaux 3.2-1 et 3.2-2 présentent les résultats pour l'ensemble des contacts hospitaliers (qu'il y ait ou non des notes cliniques) avec les résultats pour les seuls contacts avec au moins un enregistrement de notes cliniques. L'algorithme amélioré d'identification de la dépendance aux opioïdes a dénombré 1 370 827 contacts liés aux opioïdes au total. Le cinquième (20,3 %) de ces contacts était relevé exclusivement par l'élément TLN et aurait été oublié par un algorithme visant uniquement les codes médicaux. Si le traitement portait uniquement sur les contacts avec notes cliniques disponibles, une petite proportion de ces contacts (0,9 %) était relevée par le seul élément des codes médicaux (tableau 3.2-1).

L'algorithme des troubles concomitants a recensé au total 659 225 contacts liés aux opioïdes comme TUS ou PSM seulement ou comme troubles concomitants. Précisons que le seul élément TLN a relevé 10,3 % des contacts. Dans le cas des contacts avec notes cliniques disponibles, dans le cas de 10 542 contacts relevés, soit 6,0 % des cas rapportés, seule la composante code médical a permis de les relever (voir le tableau 3.2-2).

Tableau 3.2-1
Nombre et pourcentage de contacts liés aux opioïdes avec ou sans notes cliniques disponibles qui ont été relevés respectivement par les éléments codes médicaux , TLN et les deux algorithmes pour l'ensemble de données de la National Hospital Care Survey de 2016

	Tous contacts confondus		Seuls les contacts avec notes médicales	
	Nombre	Pourcentage	Nombre	Pourcentage
Élément codes seulement	1 060 495	77,4 %	2 819	0,9 %
Élément TLN seulement	277 958	20,3 %	277 958	88,8 %
Les deux	32 374	2,4 %	32 374	10,3 %
Total	1 370 827	100,0 %	313 151	100,0 %

NOTE : TLN, traitement du langage naturel.
SOURCE : National Center for Health Statistics, National Hospital Care Survey, 2016.

Tableau 3.2-2
Nombre et pourcentage de contacts TUS, PSM et troubles concomitants avec ou sans notes cliniques disponibles qui ont été relevés respectivement par les éléments codes médicaux, TLN et les deux algorithmes pour l'ensemble de données de la NHCS de 2016

	Tous contacts confondus		Seuls contacts avec notes cliniques	
	Nombre	Pourcentage	Nombre	Pourcentage
Élément codes seulement	494 458	75,0 %	10 542	6,0 %
Élément TLN seulement	67 584	10,3 %	67 584	38,6 %
Les deux	97 183	14,7 %	97 183	55,4 %
Total	659 225	100,0 %	175 309	100,0 %

NOTE : TLN, traitement du langage naturel.
SOURCE : National Center for Health Statistics, National Hospital Care Survey, 2016.

4. Analyse

4.1 Leçons tirées de la création d'une capacité en science des données

Le TLN peut constituer un précieux outil d'analyse des données hospitalières présentées même lorsque des codes médicaux normalisés sont disponibles, puisque ces codes peuvent poser un grand nombre de problèmes logistiques. Toutefois, une bonne prise en charge de formats diversifiés de données a son importance et la gestion des logiciels peut s'appuyer sur des facteurs indépendants de la volonté du chercheur, comme dans une exploitation de l'information en environnement fermé. Ajoutons que le TLN peut accroître le rappel, mais souvent au détriment de la précision.

Les éléments de mesure du rendement algorithmique dans les deux projets font ressortir l'importance d'adopter une démarche d'ensemble pour définir les concepts sanitaires d'intérêt dans les demandes administratives et les dossiers électroniques de santé en ce qui concerne les troubles liés aux opioïdes et les troubles concomitants. Nous avons constaté que les algorithmes qui combinent une recherche sur codes médicaux et les techniques d'extraction textuelle donnent globalement les meilleurs résultats et permettent de relever les contacts hospitaliers qui auraient été manqués autrement.

Les systèmes de collecte de données mis au service de la production des statistiques officielles exploitent des données de plus en plus complexes et volumineuses. Les défis que présente dans notre étude l'intégration de ces méthodes démontrent l'importance d'édifier une plateforme d'analyse de données offrant une puissance de calcul convenant à des mégadonnées, aisément adaptable à l'implantation des meilleures solutions logicielles disponibles et accessible à tous les utilisateurs désignés à l'aide d'appareils approuvés, et ce, aussi bien sur place qu'à distance. Ajoutons que, en fournissant aux répondants des outils pour extraire et présenter les données sous une forme normalisée, on se trouve à alléger le fardeau de réponse et à améliorer la qualité de l'information. Le NCHS a rédigé un guide de mise en œuvre HL7 qui indique à l'hôpital ou au fournisseur en dossiers électroniques de santé, quels sont les éléments d'information désirés et les formats privilégiés (HL7 International, 2021).

4.2 Prochaines étapes du projet

Des ensembles de données analytiques comportant des variables pour des algorithmes améliorés ont été mis à la disposition des chercheurs dans le centre de données de recherche du NCHS. On trouvera d'autres renseignements sur ces ensembles à l'adresse https://www.cdc.gov/rdc/b1datatype/dt1224h.htm. Le code machine Python de l'élément TLN sera également présenté d'ici le printemps 2022 à l'adresse https://github.com/oxf7.

Nous avons réalisé une étude de validation qui implique l'extraction de certains contacts hospitaliers pour des annotateurs externes et nous analysons actuellement les erreurs des algorithmes en fonction de ces résultats. L'algorithme sera affiné et une version à jour sera produite.

Bibliographie

Chapman W.M., Bridewell, W., Hanbury, P., Cooper, G.F., et Buchanan B.G. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 34(5):301–10. DOI: 10.1006/ jbin.2001.1029

HL7 International. HL7 CDA® R2 Implementation Guide: National Health Care Surveys (NHCS), R1 STU Release 3 – US Realm.

National Center for Health Statistics Research Data Center. (2020). Linked Data on Hospitalizations, Mortality, and Drugs: Data from the National Hospital Care Survey 2016, National Death Index 2016-2017, and the Drug Involved Mortality 2016-2017. Disponible à : https://www.cdc.gov/nchs/data/nhcs/Task-3-Doc-508.pdf.

National Center for Health Statistics Research Data Center. (2021). Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data: Capstone to an Existing FY18 PCORTF Project. Disponible à : https://www.cdc.gov/nchs/data/nhcs/FY19-RDC-2021-06-01-508.pdf.

SpaCy: Industrial-strength natural language processing (TLN) with Python and Cython. Disponible à : https://github.com/explosion/spacy-models/releases/tag/ en_core_web_md-2.2.5

White, D.G., Adams, N.B., Brown, A.M., O'Jiaku-Okorie, A., Badwe, R., Shaikh, S., et Adegboye, A. (2021b). Enhancing identification of opioid-involved health outcomes using National Hospital Care Survey data. National Center for Health Statistics Vital Health Stat 2(188). https://www.cdc.gov/nchs/data/series/sr_02/sr2-188.pdf.