

Synthèse des données probantes

Exactitude des algorithmes des bases de données administratives pour la détermination des cas de trouble du spectre de l'autisme, de trouble du déficit de l'attention/hyperactivité et de trouble du spectre de l'alcoolisation fœtale : revue systématique

Siobhan O'Donnell, M. Sc.; Sarah Palmeter, M.H.P.; Meghan Laverty, M. Sc.; Claudia Lagacé, M. Sc.

Cet article a fait l'objet d'une évaluation par les pairs.

 Diffuser cet article sur Twitter

Résumé

Introduction. L'objectif de cette revue systématique était d'évaluer la validité des algorithmes des bases de données administratives utilisés pour repérer les cas de trouble du spectre de l'autisme (TSA), de trouble du déficit de l'attention/hyperactivité (TDAH) et de trouble du spectre de l'alcoolisation fœtale (TSAF).

Méthodologie. L'équipe de recherche a interrogé les systèmes MEDLINE, Embase, Global Health et PsycInfo afin de trouver des études, publiées en anglais ou en français entre 1995 et 2021, qui valident des algorithmes servant à recenser les cas de TSA, de TDAH et de TSAF dans les bases de données administratives. L'équipe de recherche a également consulté la littérature grise et les listes de références des études incluses. Deux évaluateuses ont procédé, de manière indépendante, à la sélection de la littérature, à l'extraction de l'information pertinente, à l'évaluation de la qualité des rapports, au risque de biais et à l'applicabilité et elles ont réalisé une synthèse qualitative des données probantes. PROSPERO CRD42019146941.

Résultats. Parmi les 48 articles évalués en texte intégral, 14 ont été inclus dans la revue. Aucune étude n'a été trouvée pour le TSAF. Malgré les sources de biais possibles et la grande hétérogénéité des études, les résultats donnent à penser que l'augmentation du nombre de codes de diagnostic de TSA qui requiert une seule source de données augmente la spécificité et la valeur prédictive positive au détriment de la sensibilité. Les algorithmes les plus performants pour l'identification des cas de TSA reposent sur une combinaison de sources de données, la base de données sur les demandes de remboursement des médecins étant la meilleure source. Une étude a révélé que les données sur l'éducation pourraient améliorer l'identification des cas de TSA (sensibilité accrue) chez les enfants d'âge scolaire lorsqu'elles sont combinées aux données sur les demandes de remboursement des médecins; toutefois, d'autres études incluant des sujets sans TSA sont nécessaires pour évaluer pleinement l'exactitude diagnostique de ces algorithmes. Pour ce qui est du TDAH, il n'y a pas eu suffisamment d'information pour évaluer l'incidence du nombre de codes de diagnostic ou d'autres sources de données sur l'exactitude des algorithmes.

Conclusion. Selon certaines données probantes, il est possible de repérer les cas de TSA et de TDAH à l'aide de données administratives; toutefois, il existe peu d'études qui évaluent la capacité des algorithmes à établir une distinction fiable entre les sujets qui présentent le trouble étudié et les sujets qui ne le présentent pas. Il n'y a aucune donnée probante sur le TSAF. Des études de qualité méthodologique supérieure sont nécessaires pour comprendre tout le potentiel de l'utilisation des données administratives pour l'identification de ces troubles.

Points saillants

- Peu d'études ont validé les algorithmes des bases de données administratives servant à recenser les cas de TSA et de TDAH. Aucune étude de validation n'a été trouvée pour le TSAF.
- En raison de la grande hétérogénéité des études dans la façon dont elles ont été conçues et menées, il n'a pas été possible de faire une synthèse quantitative des résultats.
- Selon certaines données probantes, il est possible de repérer les cas de TSA et de TDAH à l'aide de données administratives; toutefois, il existe peu d'études qui évaluent la capacité des algorithmes à établir une distinction fiable entre les sujets qui présentent le trouble étudié et les sujets qui ne le présentent pas.
- Les algorithmes les plus performants utilisés pour l'identification des cas de TSA reposent sur une combinaison de sources de données administratives, les données sur les demandes de remboursement des médecins étant la meilleure source.
- Des études de qualité supérieure sont essentielles pour tirer pleinement parti des données administratives pour la surveillance de ces troubles et la recherche en la matière.

Rattachement des auteures :

Agence de la santé publique du Canada, Ottawa (Ontario), Canada

Correspondance : Siobhan O'Donnell, Centre de surveillance et de recherche appliquée, Direction générale de la promotion de la santé et de la prévention des maladies chroniques, Agence de la santé publique du Canada, 785, avenue Carling, IA 6806A, Ottawa (Ontario) K1A 0K9; tél. : 613-301-7325; courriel : siobhan.odonnell@phac-aspc.gc.ca.

Mots-clés : trouble du spectre de l'autisme, trouble du déficit de l'attention/hyperactivité, trouble du spectre de l'alcoolisation fœtale, algorithmes, étude de validation, données administratives, surveillance en santé publique

Introduction

Les troubles neurodéveloppementaux, un ensemble d'affections qui apparaissent tôt dans la vie, se caractérisent par des déficits du développement physique, de l'apprentissage, du langage ou du comportement¹. Sans effacer les vastes répercussions personnelles et sociétales qu'ils entraînent, on sait que la détection et l'intervention précoces permettent d'améliorer les résultats chez les personnes atteintes de certains types de troubles neurodéveloppementaux, notamment le trouble du spectre de l'autisme (TSA), le trouble du déficit de l'attention/hyperactivité (TDAH) et le trouble du spectre de l'alcoolisation fœtale (TSAF)²⁻⁵. Une meilleure compréhension du fardeau épidémiologique de ces troubles au Canada est essentielle à la mise en œuvre des politiques publiques, notamment pour l'établissement de programmes et de services.

Les bases de données administratives sur la population, qui sont conçues pour la gestion du système de santé et la rémunération des médecins, offrent un moyen efficace et économique de fournir des données épidémiologiques longitudinales. C'est pourquoi ces données sont de plus en plus utilisées pour assurer la surveillance des maladies chroniques⁶⁻⁸, mener des recherches sur les maladies et les résultats des traitements⁹ et réaliser des études sur la qualité des soins^{10,11}. Toutefois, les bases de données administratives sur la santé n'ont pas que des avantages : elles ont aussi certaines limites, en particulier un risque d'erreur de classification¹².

L'exactitude des codes de diagnostic ou de leur combinaison (algorithme) à des fins de surveillance ou de recherche¹³ dépend de plusieurs facteurs, dont la qualité de la base de données, l'affection particulière à identifier et la validité des codes de diagnostic au sein du groupe de patients¹². Par conséquent, des études de validation sont nécessaires pour évaluer l'exactitude des algorithmes servant à la détermination des cas¹⁴. Cette validation consiste à quantifier le nombre de fois où l'algorithme

donne le même résultat qu'un test de référence, par exemple un diagnostic au dossier médical¹². En ce sens, l'algorithme peut être considéré comme un test diagnostique et on peut effectuer des mesures de l'exactitude diagnostique. Les résultats des études de validation sont généralement présentés sous forme d'estimations de la sensibilité et de la spécificité de l'algorithme, qui expriment la mesure dans laquelle l'algorithme réussit à identifier correctement les personnes qui ont l'affection étudiée et celles qui ne l'ont pas, respectivement^{15,16}. D'autres mesures statistiques de l'exactitude diagnostique sont utilisables, par exemple la valeur prédictive positive (VPP) et la valeur prédictive négative (VPN).

À notre connaissance, aucun examen publié n'a évalué la validité des algorithmes des bases de données administratives sur la santé pour la surveillance des troubles neurodéveloppementaux que sont le TSA, le TDAH et le TSA et pour la recherche en la matière. Ainsi, l'objectif principal de cette revue systématique était de combler cette lacune. L'objectif secondaire était d'examiner l'incidence du couplage de données administratives liées à la santé et de données administratives non liées à la santé (éducation ou services sociaux) sur l'exactitude de ces algorithmes.

Méthodologie

Les résultats de cette revue systématique sont présentés conformément à la déclaration PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)¹⁷. Aucune approbation éthique n'était requise car nous n'avons pas recueilli de données primaires.

Protocole et enregistrement

Le protocole de la revue systématique a été enregistré dans le registre prospectif international des revues systématiques (PROSPERO 2019 CRD42019146941), a été publié le 16 décembre 2019 et est accessible en ligne à l'adresse https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019146941.

Stratégie de recherche

Une recherche systématique dans les systèmes MEDLINE, Embase, Global Health et PsyInfo a été effectuée afin de répertorier toutes les études de validation fondées sur des données administratives pour

déterminer les cas de TSA, de TDAH ou de TSAF et qui ont été publiées, en anglais ou en français, entre janvier 1995 et mars 2021. L'année 1995 a été choisie comme année de début pour les recherches dans les bases de données car il s'agit également de l'année de début de la collecte des données dans le Système canadien de surveillance des maladies chroniques, un réseau de collaboration composé de systèmes provinciaux et territoriaux de surveillance administrative de la santé financé par l'Agence de la santé publique du Canada. Une bibliothécaire de référence a élaboré la stratégie de recherche à l'aide de vedettes-matières médicales et de mots-clés liés aux troubles étudiés (par ex. « exp trouble du spectre de l'autisme/ »), aux données administratives (par ex. « exp assurance maladie/ »), aux tests de référence (par ex. « exp dosiers médicaux/ ») et aux tests de validation (par ex. « sensibilité et spécificité/ »). La stratégie de recherche initiale a été élaborée dans MEDLINE et examinée par des pairs avant d'être adoptée pour les autres bases de données (annexe A). De plus, la littérature grise a fait l'objet de deux types de recherche : une recherche avancée sur Google et une recherche parmi les sites Web d'organisations et d'organismes compétents. Enfin, les listes de références des rapports de surveillance pertinents tirés de la littérature grise ainsi que les articles qui répondent aux critères d'admissibilité de la revue ont fait l'objet de recherches manuelles pour trouver d'autres études.

Critères d'admissibilité

Pour être incluses, les études, quelle que soit leur conception, devaient comporter :

- l'évaluation ou la validation d'un ou de plusieurs algorithmes de bases de données administratives sur la santé par rapport à un test de référence (critères cliniques établis, diagnostic au dossier médical, dossier médical électronique ou mesure autodéclarée par le patient) pour identifier un cas de TSA, de TDAH ou de TSAF;
- au moins une mesure de l'exactitude diagnostique (sensibilité, spécificité, VPP, VPN, aire sous la courbe ROC [Receiver Operating Characteristic] [ou statistique C], indice de Youden, coefficient kappa ou rapport de vraisemblance).

Un algorithme de base de données administratives a été défini comme un ensemble

de règles servant à repérer les cas de maladie à partir de données administratives, en fonction d'éléments comme le type de source de données, le nombre d'années de données administratives, le ou les code(s) de diagnostic ou de médicament et le nombre d'enregistrements de données administratives (contacts) avec un ou plusieurs code(s) de diagnostic ou de médicament¹³. Il devait inclure des données administratives sur la santé mais pouvait également inclure d'autres types de données administratives, comme des données sur l'éducation ou les services sociaux.

Ces algorithmes pouvaient être fondés sur les données administratives provenant soit d'une base de données administratives sur la santé, soit d'un système d'information clinique ou sur la santé. Une base de données administratives sur la santé a été définie comme un ensemble de renseignements recueillis systématiquement ou passivement à des fins administratives seulement pour la gestion des soins de santé des patients¹⁸, alors qu'un système d'information clinique ou sur la santé a été défini comme un ensemble de données administratives complétées par des renseignements cliniques détaillés au moyen de dossiers de santé électroniques¹⁹.

Les résumés, les éditoriaux et les commentaires ont été exclus de cette revue, tout comme les études publiées avant 1995 ou dans une langue autre que l'anglais ou le français.

Sélection des études et collecte des données

Deux évaluateuses (CL et SO) ont procédé chacune de leur côté à un tri fondé sur le titre et le résumé de toutes les notices bibliographiques et de tous les articles issus des recherches dans les bases de données électroniques, de la littérature grise et des listes de références des rapports de surveillance pour en déterminer l'admissibilité. En l'absence de consensus sur une étude donnée, cette dernière a été soumise à un deuxième tri. Pour chaque étude retenue à l'étape du tri fondé sur le titre et le résumé, deux évaluateuses (ML et SO) ont passé en revue de manière indépendante le texte intégral de l'étude pour en déterminer l'admissibilité et, lorsque l'étude était exclue, la raison de l'exclusion a été consignée. En cas de désaccord entre les évaluateuses sur l'inclusion ou l'exclusion d'un article, une troisième évaluateuse (CL) a été consultée.

Les listes de références de tous les articles retenus à l'étape de l'examen du texte intégral ont fait l'objet d'une recherche manuelle fondée sur le même processus de tri à deux niveaux mené par deux évaluateuses (SO et SP).

Les renseignements pertinents, extraits des articles retenus à l'aide d'un modèle élaboré pour cette revue systématique et mis à l'essai avant utilisation, sont l'auteur, l'année, l'emplacement géographique, la cohorte de l'étude, le type de source de données administratives, les algorithmes des bases de données administratives et les éléments connexes, le test de référence, les critères diagnostiques de référence et les mesures de l'exactitude diagnostique. L'extraction a été réalisée par une évaluateuse (ML) et les résultats ont été vérifiés par une seconde évaluateuse (SO). Tout désaccord a été résolu par consensus ou, au besoin, par une tierce partie (CL).

Évaluations de la qualité des rapports, du risque de biais et de l'applicabilité

Les études retenues ont fait l'objet d'une évaluation de la qualité des rapports à l'aide de la liste de vérification modifiée de 40 points STARD (Standards for the Reporting of Diagnostic Accuracy Studies)¹² (annexe B) et d'une évaluation du risque de biais et de l'applicabilité à l'aide de l'outil révisé QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies)²⁰ (annexe C). Ces évaluations ont été effectuées par une évaluateuse (ML) et vérifiées par une seconde évaluateuse (SO). Tout désaccord a été résolu par consensus ou, au besoin, par une troisième évaluateuse (CL).

Synthèse et analyse des données

En raison de la grande hétérogénéité des études dans la façon dont elles ont été conçues et menées, il n'a pas été possible de réaliser une synthèse quantitative des résultats, mais l'équipe de recherche a effectué une synthèse narrative des résultats au moyen de texte et de tableaux pour les troubles étudiés. Les constats au sein d'une même étude et entre études ont été analysés conformément aux directives du Centre for Reviews and Dissemination²¹. Dans ses recommandations finales, l'équipe de recherche a pris en considération non seulement l'exactitude diagnostique des algorithmes des bases de données administratives inclus dans les études, mais aussi la qualité des rapports,

le risque de biais et les préoccupations relatives à l'applicabilité de chaque étude.

Résultats

Résultats de la recherche

Le diagramme de flux PRISMA de la figure 1 présente le processus de sélection des études¹⁷. Au total, 5918 documents ont été répertoriés dans le cadre des recherches dans les bases de données, et 11 autres documents ont été repérés dans d'autres sources (littérature grise et listes de référence des rapports de surveillance). Après l'élimination des doublons, 4133 documents issus des recherches dans les bases de données ont été soumis à un tri (d'après le titre et le résumé) et 4085 d'entre eux ont été jugés inadmissibles et ont donc été exclus. Parmi les 48 documents ayant fait l'objet d'une évaluation en texte intégral pour admissibilité, 34 ont été exclus (17 n'utilisaient pas de base de données administratives sur la santé, 6 utilisaient une base de données administratives sur la santé mais n'étaient pas des études de validation, 8 ne validaient pas d'algorithme de base de données administratives sur la santé et 3 ont été exclus pour d'autres raisons) et 14 documents (études) ont été retenus pour la revue. Aucun des 11 documents répertoriés dans la littérature grise et les listes de références des rapports de surveillance n'a été inclus dans la revue. Aucune autre étude n'a été trouvée à la suite de la recherche manuelle dans les listes de références des articles inclus.

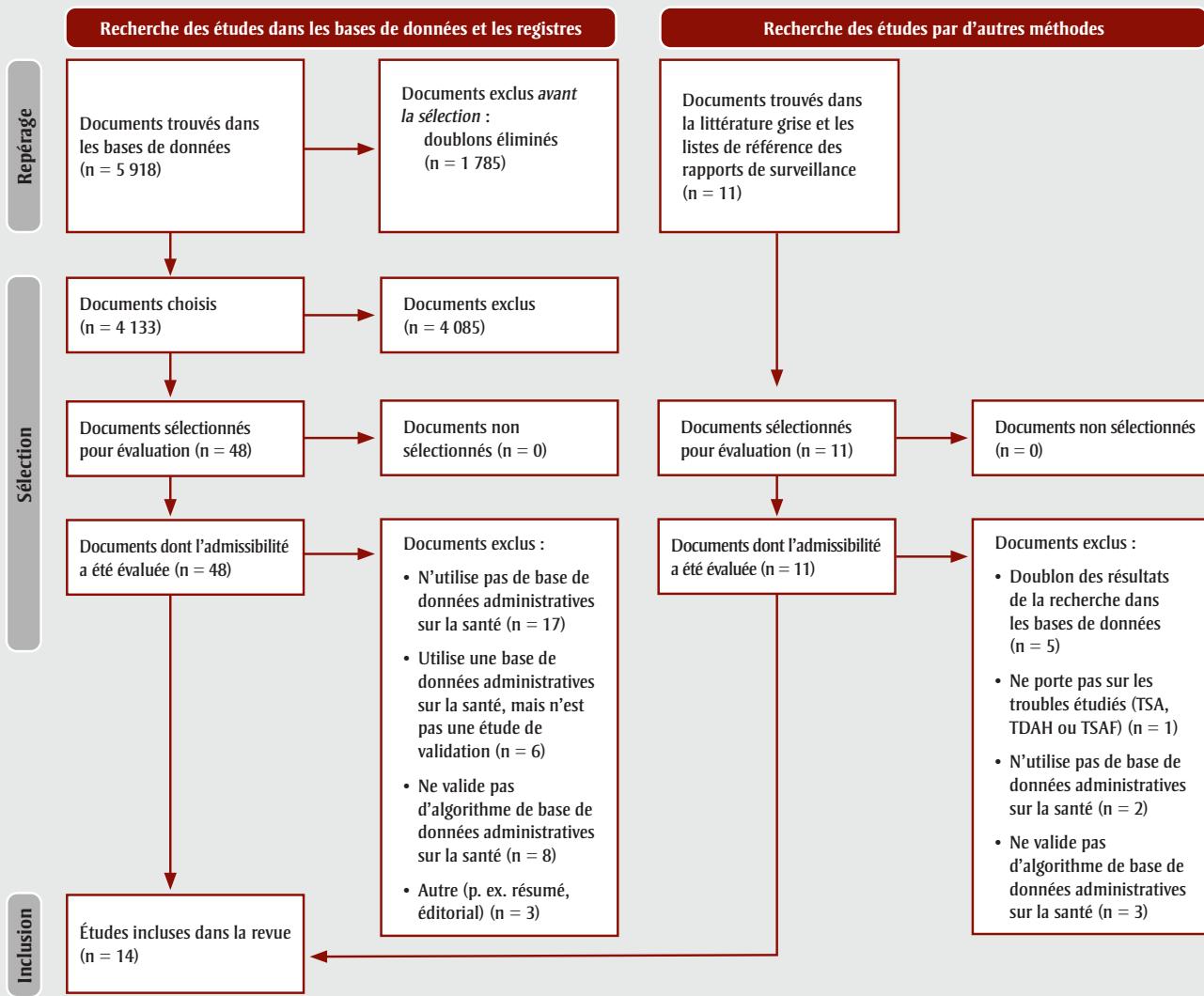
Caractéristiques des études incluses

Les caractéristiques des 14 études incluses²²⁻³⁵ sont présentées dans le tableau 1. Dix études portent sur le TSA²²⁻³¹ et les quatre autres sur le TDAH³²⁻³⁵. Aucune étude n'a été trouvée pour le TSAF.

Études sur le TSA

Plusieurs études validant des algorithmes utilisés pour recenser les cas de TSA ont été publiées entre 2009²⁸ et 2021^{23,24}. Cinq études ont été menées au Canada^{22-24,27,28}, deux aux États-Unis^{25,26}, une au Royaume-Uni²⁹, une au Danemark³⁰ et une en Norvège³¹. Ces dix études avaient pour population cible les enfants ou les jeunes²²⁻³¹, bien que la tranche d'âge étudiée ne soit mentionnée que dans sept d'entre elles^{22-27,31}.

FIGURE 1
Diagramme de flux PRISMA



Remarque : Modèle PRISMA tiré de Page MJ et al¹⁷.

La taille des échantillons des cohortes de validation variait entre 37²⁹ et 10 000^{23,24} patients. À l'origine, les patients ont été sélectionnés à partir des codes de diagnostic dans la base de données administrative pour cinq études^{25,26,29-31} et pour l'un des deux échantillons utilisés dans une étude²⁷. Seules cinq études incluaient un groupe de comparaison sans TSA^{22-25,28}. La prévalence du TSA dans la cohorte de validation variait entre 1,1 %^{23,24} et 67,9 %²².

Six études ont utilisé des bases de données administratives sur la santé^{22,23,25,28,30,31}, deux ont utilisé un système d'information clinique ou sur la santé^{24,29}, une a utilisé à la fois des bases de données administratives sur la santé et un système d'information clinique ou sur la santé²⁶ et une

autre a utilisé des bases de données administratives sur la santé combinées à une source de données sur l'éducation²⁷. La source de données la plus courante était une combinaison de données sur les patients externes et les patients hospitalisés^{22,23,28,30,31}.

Divers codes de diagnostic ont été utilisés, à savoir les codes de la Classification internationale des maladies, huitième révision (CIM-8)³⁰, les codes de la Classification internationale des maladies, neuvième révision (CIM-9)^{22,23,25-28}, les codes de la Classification internationale des maladies, dixième révision (CIM-10)^{22,23,27,28,30,31}, les codes de facturation du Régime d'assurance-maladie de l'Ontario pour les médecins^{23,24}, les codes Read²⁹ et des codes

spécifiques aux services d'éducation et de santé mentale²⁷. Le nombre d'algorithmes validés dans chaque étude variait entre 1²⁹⁻³¹ et 153²³.

Plusieurs tests de référence ont été utilisés, le plus fréquent étant un diagnostic au dossier médical^{23,24,27,29}.

La VPP a été la mesure la plus souvent employée : elle a été calculée dans neuf des dix études^{22-27,29-31}. Seulement trois études ont eu recours à au moins quatre mesures de l'exactitude diagnostique²²⁻²⁴.

Études sur le TDAH

Plusieurs études validant des algorithmes utilisés pour recenser les cas de TDAH ont été publiées entre 2014³² et 2020³⁵. Deux

TABLEAU 1
Caractéristiques des études incluses

Premier auteur, année Pays	Population de validation Âge	Taille de l'échantillon	Source de données administratives	Années de données administratives	Codes de diagnostic inclus dans l'algorithme
TSA					
Bickford, 2020 ²² Canada	Enfants de 1 à 14 ans, nés en Colombie-Britannique entre le 1 ^{er} avril 2000 et le 31 décembre 2009, évalués dans l'un des centres du British Columbia Autism Assessment Network entre le 1 ^{er} avril 2004 et le 31 décembre 2014 ou ayant reçu une désignation de TSA du ministère de l'Éducation entre le 1 ^{er} septembre 2004 et le 30 juin 2015.	8 670 (cas et non-cas)	Bases de données administratives sur la santé ^a : dossiers de congé de l'hôpital et demandes de remboursement des médecins	2000-2014	Dossiers de congé de l'hôpital : CIM-9 299.x; CIM-10-CA F84.x Demandes de remboursement des médecins : CIM-9 299.x
De 1 à 14 ans.					
Brooks, 2021 ²³ Canada	Enfants et jeunes de 1 à 24 ans au 31 décembre 2011, dont les données figurent dans la base de données Electronic Medical Record Primary Care (base de données alimentée par plus de 350 médecins de famille de l'Ontario) et dont la date de naissance est valide, qui sont inscrits auprès d'un médecin praticien en exercice utilisant le dossier médical électronique (DME) depuis plus de 2 ans, qui sont vivants à la date de chargement des données et qui ont un DME depuis au moins 1 an.	10 000 (cas et non-cas)	Bases de données administratives sur la santé ^a : dossiers de congé de l'hôpital, visites à l'urgence, chirurgies externes, demandes de remboursement des médecins	ND	Dossiers de congé de l'hôpital, chirurgies externes : CIM-9 299.x; CIM-10-CA F84.x Demandes de remboursement des médecins : code de facturation du RAMO pour les médecins 299
De 1 à 24 ans.					
Brooks, 2021 ^{24,b} Canada	Enfants et jeunes de 1 à 24 ans au 31 décembre 2011, dont les données figurent dans la base de données Electronic Medical Record Primary Care (base de données alimentée par plus de 350 médecins de famille de l'Ontario) et dont la date de naissance est valide, qui sont inscrits auprès d'un médecin praticien en exercice utilisant le dossier médical électronique (DME) depuis plus de 2 ans, qui sont vivants à la date de chargement des données et qui ont un DME depuis au moins 1 an.	10 000 (cas et non-cas)	Système d'information clinique ou sur la santé ^c : dossiers de santé électroniques	ND	Codes de facturation du RAMO pour les médecins 299, 315
De 1 à 24 ans.					

Suite à la page suivante

TABLEAU 1 (suite)
Caractéristiques des études incluses

Premier auteur, année Pays	Population de validation Âge	Taille de l'échantillon	Source de données administratives	Années de données administratives	Codes de diagnostic inclus dans l'algorithme
Burke, 2014 ²⁵ États-Unis	Enfants et jeunes de 2 à 20 ans au moment de la première réclamation pour un TSA ou pour un trouble lié au TSA, qui sont assurés par un important régime national privé d'assurance-maladie. Pour être admissibles, les cas devaient être inscrits au régime pendant une période minimale continue de 6 mois avant et après la première réclamation pour un TSA ou pour un trouble lié au TSA et n'avoir aucune réclamation associée à un diagnostic de trouble désintégratif de l'enfance ou de syndrome de Rett.	432 (cas et non-cas)	Base de données administratives sur la santé ^a : réclamations d'assurance privée pour des services médicaux, pharmaceutiques et de santé comportementale	2001-2009	TSA : CIM-9 299.00 à 299.01, 299.80 à 299.81, 299.9 (peu importe la position) Troubles liés au TSA : CIM-9 317.00, 318.00, 318.10, 318.20, 319.00, 759.50, 759.83, 771.00, 348.30, 348.80, 348.90, 783.42, V79.80, V79.90, 315.30, 315.31, 313.32, 315.40, 315.50, 315.80, 315.90, 330.8, 299.1
Coleman, 2015 ²⁶ États-Unis	De 2 à 20 ans. Enfants et jeunes de moins de 18 ans inscrits à l'un des régimes d'assurance-maladie participants en date de décembre 2010 et à qui avait été attribué au moins un code de diagnostic de TSA, mais qui n'avaient pas été diagnostiqués dans un centre spécialisé en TSA.	1 272 (cas seulement)	Base de données administratives sur la santé ^a et système d'information clinique ou sur la santé ^c : réclamations d'assurance et dossiers de santé électroniques	1995-2010	CIM-9 299.0, 299.9, 299.8
Coo, 2017 ²⁷ Canada	Moins de 18 ans. Enfants de 2 à 14 ans, nés entre 1997 et 2009, qui ont un diagnostic administratif de TSA ou dont le cas a été confirmé par un fournisseur de services en matière de comportement ou de handicap pour les enfants et les jeunes du Manitoba le 31 décembre 2011 ou avant cette date.	2 610 (cas seulement)	Bases de données administratives sur la santé ^a et base de données sur l'éducation : dossiers de congé de l'hôpital, demandes de remboursement des médecins, services de santé mentale et données sur l'éducation	1997-2011	Dossiers de congé de l'hôpital : CIM-9-MC 299.0, 299.8, 299.9; CIM-10-CA F84.0, F84.1, F84.5, F84.8, F84.9 (dans tout champ de diagnostic) Demandes de remboursement des médecins : CIM-9-MC 299.x (diagnostic « principal ») Données sur l'éducation : CATEGORYN=ASD (enfant ayant reçu du financement au titre de la catégorie des besoins particuliers pour le TSA) Services de santé mentale : NDC-A 312 0.92 (inscription à un programme de traitement de l'autisme)

Suite à la page suivante

TABLEAU 1 (suite)
Caractéristiques des études incluses

Premier auteur, année Pays	Population de validation Âge	Taille de l'échantillon	Source de données administratives	Années de données administratives	Codes de diagnostic inclus dans l'algorithme
Dodds, 2009 ²⁸ Canada	Enfants nés entre 1989 et 2002, et évalués pour un TSA par une équipe de spécialistes entre 2001 et 2005. Âge non indiqué.	264 (cas et non-cas)	Bases de données administratives sur la santé ^a : dossiers de congé de l'hôpital, demandes de remboursement des médecins et données sur les patients externes en santé mentale	1989-2005	CIM-9 299.x ou CIM-10 F84.x (champ de diagnostic principal ou secondaire)
Hagberg, 2017 ²⁹ Royaume-Uni	Enfants issus d'un accouchement simple nés entre 1990 et 2011, ayant fait l'objet d'un suivi d'au moins trois ans depuis la naissance. Âge non indiqué.	37 (cas seulement)	Système d'information clinique ou sur la santé ^c : dossiers de santé électroniques	1990-2014	Codes Read : E140.00, E140000, E140100, E140.12, E140.13, E140z00, Eu84000, Eu84011, Eu84012, Eu84100, Eu84z11, Eu84500, Eu84.00, Eu84y00, Eu84z00
Lauritsen, 2010 ³⁰ Danemark	Enfants nés entre 1990 et 1999 dont le ou les parent(s) ou tuteur(s) légaux résident au Danemark et qui ont reçu un diagnostic d'autisme infantile déclaré. Âge non indiqué.	499 (cas seulement)	Base de données administratives sur la santé ^a : données sur les patients hospitalisés et les patients externes en psychiatrie	1990-2001	CIM-8 299.00 ou CIM-10 F84.0 (diagnostic principal ou secondaire)
Surén, 2019 ³¹ Norvège	Enfants nés entre 1999 et 2009, inscrits à l'étude de cohorte Norwegian Mother, Father and Child Cohort Study, qui ont reçu un diagnostic d'autisme déclaré dans le registre des patients norvégien entre 2008 et 2014, qui étaient âgés de 5 à 15 ans à la fin du suivi, dont le dossier de patient était disponible et qui n'avaient pas fait l'objet d'une évaluation clinique dans le cadre de l'étude sur l'autisme. De 5 à 15 ans.	553 (cas seulement)	Base de données administratives sur la santé ^a : données des fournisseurs de soins de santé mentale, des hôpitaux somatiques et des experts-conseils du secteur privé	2008-2014	CIM-10 F84.x
TDAH					
Daley, 2014 ³² États-Unis	Enfants de 3 à 9 ans au moment du premier diagnostic, assurés par l'un des huit organismes de gestion des soins ou ayant consulté l'un des deux établissements de santé communautaires entre 2004 et 2010, qui répondaient à la définition de cas de TDAH et qui n'avaient pas reçu de diagnostic de retard mental ou de trouble envahissant du développement. De 3 à 9 ans.	500 (cas seulement)	Système d'information clinique ou sur la santé ^c : dossiers de santé électroniques	2004-2010	CIM-9-MC 314.0x

Suite à la page suivante

TABLEAU 1 (suite)
Caractéristiques des études incluses

Premier auteur, année Pays	Population de validation Âge	Taille de l'échantillon	Source de données administratives	Années de données administratives	Codes de diagnostic inclus dans l'algorithme
Gruschow, 2016 ³³ États-Unis	Patients du réseau de soins de santé du Children's Hospital of Philadelphia, nés entre 1987 et 1995 (âge médian de 17,9 ans), cumulant au moins 2 visites et résidant au New Jersey au moment de leur dernière visite, qui n'avaient pas été désignés comme ayant une déficience intellectuelle et qui avaient au moins 12 ans lors de leur dernière visite. Les enfants ayant un diagnostic de TDAH consigné dans leur dossier de santé électronique et les enfants sans ce diagnostic ont été identifiés. Âge médian (intervalle interquartile) : 17,9 (15,9 à 19,1) ans	2 030 (cas) 807 (non-cas)	Système d'information clinique ou sur la santé ^c : dossiers de santé électroniques	2001+	CIM-9-MC 314.x
Mohr-Jensen, 2016 ³⁴ Danemark	Enfants et jeunes de 4 à 15 ans ayant reçu un diagnostic déclaré de trouble hyperkinétique pour la première fois entre 1995 et 2005.	372 (cas seulement)	Base de données administratives sur la santé ^a : données des hôpitaux psychiatriques	1995-2005	CIM-10 F90.x
Morkem, 2020 ³⁵ Canada	Enfants et adultes de 4 ans et plus identifiés à partir d'une seule clinique, avec une entrée valide pour l'année de naissance et le sexe, et une consultation en soins primaires au cours de l'année d'étude ou de l'année précédente (de 2008 à 2015). Les patients présentant certains problèmes médicaux ont été exclus.	246 (cas) 246 (non-cas)	Système d'information clinique ou sur la santé ^c : dossiers de santé électroniques	ND	CIM-9 314.x Ordonnances de médicaments liés au TDAH

4 ans et plus.

Abréviations : CIM-8, Classification internationale des maladies, huitième révision; CIM-9, Classification internationale des maladies, neuvième révision; CIM-9-MC, Classification internationale des maladies, neuvième révision, modification clinique; CIM-10, Classification internationale des maladies, dixième révision; CIM-10-CA, Classification internationale des maladies, dixième révision, Canada; DME, dossier médical électronique; ND, non déclaré; RAMO, Régime d'assurance-maladie de l'Ontario; TDAH, trouble du déficit de l'attention/hyperactivité; TSA, trouble du spectre de l'autisme.

^a Une base de données administratives sur la santé est définie comme un ensemble de renseignements recueillis passivement, souvent par le gouvernement et les fournisseurs de soins de santé, dans le but de gérer les soins de santé des patients (par ex. les données sur les demandes de remboursement).

^b Cette étude a également mis à l'essai des algorithmes qui comprenaient des renseignements d'identification des cas tirés d'une recherche par mot-clé du profil cumulatif des patients dans le dossier de santé électronique; toutefois, ces algorithmes n'ont pas été inclus car ils ne répondent pas à la définition d'un algorithme de base de données administratives utilisée dans le cadre de la revue.

^c Un système d'information clinique ou sur la santé est défini comme un ensemble de données administratives qui intègre les renseignements issus des dossiers de santé électroniques ou un ensemble de données administratives complétées par des renseignements cliniques détaillés.

études ont été menées aux États-Unis^{32,33}, une au Canada³⁵ et une au Danemark³⁴. Trois études avaient pour population cible les enfants ou les jeunes³²⁻³⁴, deux d'entre elles précisant la tranche d'âge^{32,34}. Une étude portait sur les adultes et les enfants âgés de quatre ans et plus³⁵.

La taille des échantillons des cohortes de validation variait entre 372³⁴ et 2 837³³ patients. À l'origine, les patients ont été sélectionnés à partir des codes de

diagnostic dans la source de données administratives pour trois études³²⁻³⁴ et à partir des codes de diagnostic et des ordonnances de médicaments dans la source de données administratives pour une étude³⁵. Seulement deux études incluaient un groupe de comparaison sans TDAH^{33,35}. La prévalence du TDAH dans la cohorte de validation variait entre 50,0 %³⁵ et 56,7 %³³.

Une étude a utilisé une base de données administratives sur la santé comprenant

des données sur les patients hospitalisés et les patients externes en psychiatrie³⁴, et trois études ont utilisé un système d'information clinique ou sur la santé, à savoir des dossiers de santé électroniques^{32,33,35}.

Deux études ont utilisé les codes de la CIM-9^{32,33}, une a utilisé les codes de la CIM-10³⁴ et une autre a utilisé les codes de la CIM-9 et des ordonnances de médicaments³⁵. Chaque étude validait un seul algorithme³²⁻³⁵. Une étude était fondée sur

les cas incidents plutôt que sur les cas prévalents de TDAH³².

Divers tests de référence ont été utilisés. Une étude a eu recours aux critères de classification clinique indiqués dans le dossier médical³⁴. Une autre étude a utilisé un diagnostic de TDAH au dossier médical³⁵. Une étude a utilisé une définition de cas clinique qui exigeait une combinaison de données tirées du dossier de santé électronique et, en l'absence de ces données, un examen manuel du dossier de santé électronique³³. Enfin, une étude a employé une combinaison de critères de classification clinique, de diagnostic au dossier médical et de liste de vérification normalisée figurant dans le dossier médical³².

La VPP a été calculée dans les quatre études³²⁻³⁵ et une seule étude a utilisé au moins quatre mesures de l'exactitude diagnostique³³.

Qualité des rapports des études incluses

Le tableau 2 présente le nombre et le pourcentage d'études incluses qui satisfont aux critères de la liste de vérification modifiée STARD utilisée pour la validation des données administratives sur la santé¹². La qualité des rapports était variable. Les domaines où la qualité des rapports était particulièrement sous-optimale, c'est-à-dire où moins de la moitié des études satisfaisaient au critère, sont répertoriés ci-dessous. Tous les détails sur les résultats relatifs à la qualité des rapports pour chaque étude retenue sont disponibles dans l'annexe B.

Études sur le TSA

En ce qui concerne les méthodes utilisées, aucune des études sur le TSA ne décrivait la gravité de l'état des patients, une seule étude revalidait les algorithmes à l'aide d'une cohorte distincte²³ et, parmi les huit études dans le cadre desquelles le test de référence était évalué, il n'y en avait que trois où les évaluateurs ignoraient la classification des patients en fonction des données administratives^{23,24,27}. En matière de résultats, seulement quatre études fournissaient un diagramme de flux^{23-25,31}, aucune ne faisait état des résultats des tests selon la gravité de l'état, seulement trois présentaient au moins quatre mesures de l'exactitude diagnostique²²⁻²⁴, seulement quatre rendaient compte de l'exactitude diagnostique par sous-groupe d'intérêt²⁵⁻²⁸ et seulement deux des neuf études qui précisaiient la VPP ou la VPN

présentaient, pour la cohorte de validation, un ratio cas-témoins qui se rapprochait de la prévalence du TSA dans la population^{23,24}.

Études sur le TDAH

En ce qui concerne les méthodes utilisées, aucune des études sur le TDAH ne décrivait la gravité de l'état des patients, aucune étude ne revalidait les algorithmes à l'aide d'une cohorte distincte et une seule étude faisait appel à des évaluateurs du test de référence qui ignoraient la classification des patients selon les données administratives³⁵. En matière de résultats, aucune étude ne faisait état des résultats des tests selon la gravité de l'état, une seule étude présentait au moins quatre mesures de l'exactitude diagnostique³³, une seule étude rendait compte de l'exactitude diagnostique par sous-groupe d'intérêt³² et aucune étude ne présentait, pour la cohorte de validation, un ratio cas-témoins qui se rapprochait de la prévalence du TDAH dans la population.

Risque de biais et préoccupations relatives à l'applicabilité des études incluses

La figure 2 présente un aperçu du risque de biais et des préoccupations relatives à l'applicabilité des études incluses par domaine de l'échelle QUADAS-2²⁰. Les évaluations ont révélé un risque de biais « élevé » ou « incertain » dans les domaines de la sélection des patients, du test de référence ainsi que du suivi et de l'organisation temporelle dans cinq ou plus des quatorze études. Toutes les études présentaient un faible risque de biais dans le domaine du test évalué en raison de l'objectivité des algorithmes des bases de données administratives. Il n'y avait aucune préoccupation relative à l'applicabilité des études à la question de la revue dans les trois domaines (sélection des patients, test évalué et test de référence). Les évaluations complètes du risque de biais et de l'applicabilité de chaque étude incluse sont disponibles dans l'annexe C.

Études sur le TSA

Sélection des patients

Trois études présentaient un risque de biais élevé^{22,25,31}, une présentait un risque de biais élevé dans l'un de ses deux échantillons²⁷ et une autre présentait un risque de biais incertain²⁶. Ces cotes de risque ont été attribuées en raison de la méthode d'échantillonnage^{25,27}, d'un manque d'information²⁶, de l'utilisation

d'un modèle cas-témoin²² ou d'exclusions inappropriées³¹.

Test de référence

Deux études présentaient un risque de biais incertain^{25,29} et une étude présentait un risque de biais incertain dans l'un de ses deux échantillons²⁷. Ces cotes de risque ont été attribuées en raison d'un manque d'information sur la rigueur du test de référence^{27,29}, du fait qu'il n'était pas précisé si les évaluateurs connaissaient ou non les résultats de l'algorithme²⁵ ou du fait que le test de référence reposait en partie sur le diagnostic déclaré par les parents²⁷.

Suivi et organisation temporelle

Cinq études présentaient un risque de biais élevé^{26-28,30,31} car les patients n'avaient pas tous été inclus dans l'analyse^{26,28,30,31} ou n'avaient pas tous été évalués au moyen du même test de référence²⁷.

Études sur le TDAH

Sélection des patients

Trois études présentaient un risque de biais élevé en raison d'exclusions inappropriées^{32,33,35}.

Test de référence

Une étude présentait un risque de biais élevé, car il était peu probable que le test de référence classe les cas correctement et les évaluateurs connaissaient les résultats de l'algorithme³³, et une étude présentait un risque de biais incertain en raison d'un manque d'information³⁵.

Suivi et organisation temporelle

Deux études présentaient un risque de biais élevé^{33,34} car les patients n'avaient pas tous été inclus dans l'analyse³⁴ ou n'avaient pas tous été évalués au moyen du même test de référence³³.

Exactitude diagnostique des algorithmes des bases de données administratives

Vu l'hétérogénéité constatée dans la façon dont les études incluses ont été conçues et menées, la synthèse présentée ci-dessous met en évidence les résultats sur l'exactitude diagnostique des algorithmes mis à l'essai dans le cadre de chaque étude plutôt qu'entre elles. Les estimations de l'exactitude diagnostique des algorithmes étaient très variables d'une étude à l'autre, probablement en raison de l'hétérogénéité observée entre les études. Cette hétérogénéité peut s'expliquer par les différences dans la méthode de sélection initiale des cas, les sources de

TABLEAU 2
Nombre et pourcentage d'études incluses qui satisfont aux critères de la liste de vérification modifiée STARD^a
pour validation des données administratives sur la santé

Section, sujet et point	Fréquence (%)	
	Études sur le TSA ^b	Études sur le TDAH ^c
TITRE, MOTS-CLÉS, RÉSUMÉ		
1. Identifie l'article comme une étude sur l'exactitude diagnostique?	10 (100)	4 (100)
2. Identifie l'article comme une étude sur des données administratives?	8 (80)	3 (75)
INTRODUCTION		
3. Indique que l'identification et la validation de la maladie font partie des objectifs de l'étude?	10 (100)	4 (100)
MÉTHODOLOGIE		
<i>Participants de la cohorte de validation</i>		
4. Décrit la cohorte de validation (cohorte de patients à laquelle le test de référence a été appliqué)?	10 (100)	4 (100)
4a. Âge?	10 (100)	4 (100)
4b. Maladie?	10 (100)	4 (100)
4c. Gravité?	0 (0)	0 (0)
4d. Lieu/région?	7 (70)	2 (50)
5. Décrit la procédure de recrutement de la cohorte de validation?	10 (100)	4 (100)
5a. Critères d'inclusion?	10 (100)	4 (100)
5b. Critères d'exclusion?	5 (50)	4 (100)
6. Décrit l'échantillonnage des patients (aléatoire, consécutif, tous, etc.)?	9 (90)	4 (100)
7. Décrit la collecte des données? (n = 8 études sur le TSA)	8 (100)	4 (100)
7a. Qui a identifié les patients et a veillé à ce que leur sélection respecte les critères de recrutement des patients? (n = 8 études sur le TSA)	8 (100)	4 (100)
7b. Qui a recueilli les données? (n = 8 études sur le TSA)	8 (100)	4 (100)
7c. Formulaire de collecte de données a priori? (n = 8 études sur le TSA)	5 (62,5)	2 (50)
7d. Comment la maladie a-t-elle été classée?	10 (100)	3 (75)
8. Y avait-il un échantillon fractionné (c.-à-d. revalidation à l'aide d'une cohorte distincte)?	1 (10)	0 (0)
<i>Méthodes de tests</i>		
9. Décrit le nombre de personnes ayant fait la lecture du test de référence, ainsi que la formation et l'expertise de ces personnes? (n = 8 études sur le TSA)	6 (75)	3 (75)
10. Si plus d'une personne a fait la lecture du test de référence, indique la mesure de la cohérence (p. ex. kappa)? (n = 6 études sur le TSA; n = 3 études sur le TDAH)	3 (50)	2 (66,7)
11. Les personnes ayant fait la lecture du test de référence (validation) ignoraient-elles les résultats de la classification des patients selon les données administratives? (p. ex. la personne qui a évalué les dossiers ignorait-elle les codes de facturation attribués dans les dossiers?) (n = 8 études sur le TSA)	3 (37,5)	1 (25)
<i>Méthodes statistiques</i>		
12. Décrit les méthodes de calcul ou de comparaison de l'exactitude diagnostique?	10 (100)	3 (75)
RÉSULTATS		
<i>Participants</i>		
13. Mentionne la période durant laquelle l'étude s'est déroulée et les dates de début et de fin du recrutement?	8 (80)	2 (50)
14. Décrit le nombre de personnes répondant aux critères d'inclusion ou d'exclusion?	10 (100)	4 (100)
15. Diagramme de flux de l'étude?	4 (40)	3 (75)
<i>Résultats des tests</i>		
16. Indique la répartition des différents degrés de gravité de la maladie?	0 (0)	0 (0)
17. Contient un tableau croisé comparant les résultats des tests étudiés aux résultats du test de référence?	9 (90)	3 (75)

Suite à la page suivante

TABLEAU 2 (suite)
Nombre et pourcentage d'études incluses qui satisfont aux critères de la liste de vérification modifiée STARD^a
pour validation des données administratives sur la santé

Section, sujet et point	Fréquence (%)	
	Études sur le TSA ^b	Études sur le TDAH ^c
Estimations		
18. Présente au moins quatre estimations de l'exactitude diagnostique? (estimations présentées dans les études incluses)	3 (30)	1 (25)
18a. Sensibilité	5 (50)	1 (25)
18b. Spécificité	4 (40)	1 (25)
18c. VPP	9 (90)	4 (100)
18d. VPN	4 (40)	2 (50)
18e. Rapports de vraisemblance	0 (0)	0 (0)
18f. Kappa	1 (10)	1 (25)
18g. Aire sous la courbe ROC / statistique C	2 (20)	0 (0)
18h. Exactitude/concordance	0 (0)	1 (25)
19. Les données sur l'exactitude sont-elles présentées pour des sous-groupes (p. ex. âge, situation géographique, sexe)?	4 (40)	1 (25)
20. Si la VPP ou la VPN est calculée, est-ce que le ratio cas-témoins de la cohorte de validation se rapproche de la prévalence de la maladie dans la population? (n = 9 études sur le TSA)	2 (22,2)	0 (0)
21. Calcule l'IC à 95 % pour chaque mesure de l'exactitude diagnostique?	6 (60)	3 (75)
ANALYSE		
22. Traite de l'applicabilité des résultats?	10 (100)	4 (100)

Abréviations : IC, intervalle de confiance; ROC, Receiver Operating Characteristic; TDAH, trouble du déficit de l'attention/hyperactivité; TSA, trouble du spectre de l'autisme; VPN, valeur prédictive négative; VPP, valeur prédictive positive.

Remarque : Les critères de la liste de vérification modifiée STARD proviennent de Benchimol et collab¹².

^a La liste de vérification STARD (Standards for the Reporting of Diagnostic Accuracy Studies) est une méthode d'évaluation de la qualité des rapports des études de validation faisant appel à des données administratives.

^b n = 10 sauf indication contraire.

^c n = 4 sauf indication contraire.

données administratives, les tests de référence et les algorithmes mis à l'essai. Par exemple, deux études^{23,24} ayant recours à la même cohorte de validation et à des algorithmes similaires ont utilisé des sources de données administratives différentes (base de données administratives sur la santé et système d'information clinique ou sur la santé) et ont fait état de mesures de rendement très différentes, à savoir la sensibilité et la VPP. Le tableau 3 présente l'exactitude diagnostique des algorithmes validés dans chaque étude incluse.

Études sur le TSA

Pour les études sur le TSA, l'exactitude diagnostique des algorithmes mis à l'essai a été résumée de trois façons.

Selon l'algorithme de base de données administratives sur la santé

Sept études ont mis à l'essai et comparé plusieurs algorithmes, chacun nécessitant plus ou moins de codes de diagnostic provenant d'une source particulière de

données administratives sur la santé (demandes de remboursement des médecins) sur une période comparable²²⁻²⁸. En général, ces études ont révélé que l'augmentation du nombre de diagnostics de TSA requis à partir des demandes de remboursement des médecins augmentait la spécificité et la VPP de l'algorithme, au détriment de la sensibilité. Par exemple, une étude a montré une sensibilité de 62,5 % et une spécificité de 83,0 % lorsque l'on utilisait un algorithme qui exigeait au moins un code de TSA provenant de la base de données sur les congés d'hôpital ou de la base de données sur les demandes de remboursement des médecins²⁸. Toutefois, lorsque le même algorithme exigeait au moins deux codes de TSA provenant de la base de données sur les demandes de remboursement des médecins, la spécificité s'améliorait (93,2 %) au prix d'une réduction spectaculaire de la sensibilité (36,9 %).

Trois études ont mis à l'épreuve la valeur d'autres sources de données administratives

sur la santé dans leurs algorithmes^{22,23,28}. Dans deux de ces études, il n'y avait pas d'amélioration significative de l'exactitude diagnostique des algorithmes qui exigeaient des codes de diagnostic à partir d'une combinaison des dossiers de congé de l'hôpital ou des demandes de remboursement des médecins (avec ou sans visite à l'urgence ou chirurgie ambulatoire) comparativement aux demandes de remboursement des médecins seulement^{22,23}. Dans une étude, l'utilisation d'au moins un code de diagnostic provenant d'une des trois sources de données (dossiers de congé de l'hôpital, données sur les consultations externes en santé mentale ou demandes de remboursement des médecins) a augmenté la sensibilité de l'algorithme de 9,6 %, au détriment de la spécificité (diminution de 7,9 %), comparativement à l'utilisation des demandes de remboursement des médecins seulement²⁸. De plus, après avoir évalué séparément l'exactitude des algorithmes en fonction de ces trois sources de données (demandes de remboursement des médecins

FIGURE 2
Risque de biais et préoccupations relatives à l'applicabilité des études incluses par domaine de l'échelle QUADAS-2^a

Étude	Risque de biais				Préoccupations relatives à l'applicabilité		
	Sélection des patients	Test évalué	Test de référence	Suivi et organisation temporelle	Sélection des patients	Test évalué	Test de référence
TSA							
Bickford, 2020 ²²	😊	😊	😊	😊	😊	😊	😊
Brooks, 2021 ²³	😊	😊	😊	😊	😊	😊	😊
Brooks, 2021 ²⁴	😊	😊	😊	😊	😊	😊	😊
Burke, 2014 ²⁵	😊	😊	?	😊	😊	😊	😊
Coleman, 2015 ²⁶	?	😊	😊	😊	😊	😊	😊
Coo, 2017 ²⁷	😊 ^b 😊 ^c	😊	?	😊	😊	😊	😊
Dodds, 2009 ²⁸	😊	😊	😊	😊	😊	😊	😊
Hagberg, 2017 ²⁹	😊	😊	?	😊	😊	😊	😊
Lauritsen, 2010 ³⁰	😊	😊	😊	😊	😊	😊	😊
Surén, 2019 ³¹	😊	😊	😊	😊	😊	😊	😊
TDAH							
Daley, 2014 ³²	😊	😊	😊	😊	😊	😊	😊
Gruschow, 2016 ³³	😊	😊	😊	😊	😊	😊	😊
Mohr-Jensen, 2016 ³⁴	😊	😊	😊	😊	😊	😊	😊
Morkem, 2020 ³⁵	😊	😊	?	😊	😊	😊	😊
😊 Risque faible		😊 Risque élevé		? Risque incertain			

Abréviations : TDAH, trouble du déficit de l'attention/hyperactivité; TSA, trouble du spectre de l'autisme.

^a Outil révisé QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) : méthode d'évaluation du risque de biais et de l'applicabilité des études diagnostiques dans quatre domaines d'étude clés : pour chaque domaine, on attribue une cote au risque de biais (faible, élevé ou incertain) et à l'applicabilité de l'étude à la question de recherche (faible, élevé ou incertain).

^b Risque de biais pour la partie de l'étude qui évalue la sensibilité à l'aide de la « cohorte de sensibilité » de l'étude.

^c Risque de biais pour la partie de l'étude qui évalue la valeur prédictive positive (VPP) en fonction des enfants ayant reçu un diagnostic administratif de TSA.

seulement, dossiers de congé de l'hôpital seulement et données sur les consultations externes en santé mentale seulement), la même étude a révélé que les codes de diagnostic de TSA tirés des demandes de remboursement des médecins donnaient lieu à l'algorithme le plus performant.

Deux études ont évalué des algorithmes en faisant varier le nombre d'années prises en compte dans la recherche des codes de diagnostic de TSA provenant des demandes de remboursement des médecins (par ex. deux codes ou plus sur une période de deux ans par rapport à deux codes ou plus sur une période de trois ans)^{23,24}. Ces deux études ont constaté que le fait d'augmenter le nombre

d'années prises en compte pour la recherche des codes ne se traduisait pas par une amélioration significative de l'exactitude diagnostique.

Selon le test de référence

Parmi les dix études retenues, deux faisaient varier le degré de sévérité des critères diagnostiques utilisés pour confirmer les cas de TSA^{25,26}. Ces deux études ont révélé que, lorsque l'on appliquait des exigences plus rigoureuses relativement aux éléments d'information à l'appui d'un TSA devant figurer au dossier médical, la VPP augmentait considérablement. Par exemple, dans une de ces études, la VPP est passée de 27 % à 72 % pour un algorithme nécessitant au moins un code de TSA, et elle est passée de 36 % à 87 %

pour un algorithme nécessitant au moins deux codes de TSA²⁶.

Par la combinaison des données administratives sur l'éducation et la santé

Une seule étude a validé des algorithmes faisant appel à des données administratives sur l'éducation et la santé pour repérer les cas de TSA²⁷. En général, les algorithmes qui combinaient des données sur l'éducation et des données issues des demandes de remboursement des médecins étaient associés à une amélioration de la sensibilité; toutefois, la VPP demeurait inchangée ou diminuait légèrement comparativement aux algorithmes fondés uniquement sur les données liées aux demandes de remboursement des médecins. Par exemple, dans le groupe des 6 à

TABLEAU 3
Exactitude diagnostique des études incluses

Premier auteur, année Pays	Test de référence	Algorithme(s) des bases de données administratives	Mesures de l'exactitude diagnostique (IC à 95 %) ^a
TSA			
Bickford, 2020 ²² Canada	Diagnostic clinique : Données cliniques sur le statut de TSA provenant du British Columbia Autism Assessment Network ou du ministère de l'Éducation. Tous les diagnostics posés selon une approche normalisée fondée sur les critères du DSM, à partir de l'évaluation directe de l'enfant, des renseignements fournis par la famille et de toute autre information pertinente. Les diagnostics ont été posés par des cliniciens à l'aide de l'ADOS (Autism Diagnostic Observation Schedule) et de l'ADI-R (Autism Diagnostic Interview-Revised).	≥ 1 congé de l'hôpital ou ≥ 1 demande de remboursement d'un médecin	SENS : 75 % (74 % à 76 %), SPEC : 67 % (65 % à 69 %), VPP : 82,7 %, VPN : 55,6 %, stat C : 0,71 (0,70 à 0,72), kappa : 0,40 (0,38 à 0,42)
		≥ 1 demande de remboursement d'un médecin	SENS : 74 % (72 % à 75 %), SPEC : 68 % (66 % à 69 %), VPP : 82,7 %, VPN : 54,7 %, stat C : 0,71 (0,70 à 0,72), kappa : 0,39 (0,37 à 0,41)
		≥ 1 congé de l'hôpital ou ≥ 2 demandes de remboursement d'un omnipraticien ou ≥ 1 demande de remboursement d'un pédiatre ou ≥ 1 demande de remboursement d'un psychiatre ou d'un neurologue, ou ≥ 1 demande de remboursement d'un autre spécialiste	SENS : 67 % (66 % à 69 %), SPEC : 71 % (69 % à 73 %), VPP : 83,1 %, VPN : 50,8 %, stat C : 0,69 (0,68 à 0,70), kappa : 0,35 (0,33 à 0,37)
		≥ 1 congé de l'hôpital ou ≥ 3 demandes de remboursement d'un omnipraticien ou ≥ 1 demande de remboursement d'un pédiatre ou ≥ 1 demande de remboursement d'un psychiatre ou d'un neurologue, ou ≥ 1 demande de remboursement d'un autre spécialiste	SENS : 64 % (63 % à 66 %), SPEC : 73 % (71 % à 74 %), VPP : 83,2 %, VPN : 49,1 %, stat C : 0,68 (0,67 à 0,69), kappa : 0,33 (0,31 à 0,35)
		≥ 1 congé de l'hôpital ou ≥ 2 demandes de remboursement d'un médecin	SENS : 57 % (56 % à 58 %), SPEC : 84 % (83 % à 86 %), VPP : 88,4 %, VPN : 48,2 %, stat C : 0,71 (0,70 à 0,72), kappa : 0,35 (0,33 à 0,36)
		≥ 2 demandes de remboursement d'un médecin	SENS : 55 % (54 % à 57 %), SPEC : 85 % (83 % à 86 %), VPP : 88,5 %, VPN : 47,3 %, stat C : 0,70 (0,69 à 0,71), kappa : 0,33 (0,32 à 0,35)
		≥ 1 demande de remboursement d'un pédiatre	SENS : 54 % (53 % à 56 %), SPEC : 76 % (75 % à 78 %), VPP : 82,8 %, VPN : 44,2 %, stat C : 0,65 (0,64 à 0,66), kappa : 0,26 (0,24 à 0,28)
		≥ 1 congé de l'hôpital ou ≥ 2 demandes de remboursement d'un omnipraticien ou ≥ 2 demandes de remboursement d'un pédiatre ou ≥ 2 demandes de remboursement d'un psychiatre ou d'un neurologue, ou ≥ 2 demandes de remboursement d'un autre spécialiste	SENS : 52 % (51 % à 54 %), SPEC : 86 % (85 % à 87 %), VPP : 88,7 %, VPN : 46,1 %, stat C : 0,69 (0,68 à 0,70), kappa : 0,31 (0,30 à 0,33)
		≥ 2 demandes de remboursement d'un omnipraticien ou ≥ 2 demandes de remboursement d'un pédiatre ou ≥ 2 demandes de remboursement d'un psychiatre ou d'un neurologue, ou ≥ 2 demandes de remboursement d'un autre spécialiste	SENS : 50 % (49 % à 52 %), SPEC : 87 % (85 % à 88 %), VPP : 88,8 %, VPN : 45,2 %, stat C : 0,68 (0,68 à 0,69), kappa : 0,30 (0,28 à 0,31)
		≥ 1 demande de remboursement d'un omnipraticien	SENS : 44 % (42 % à 45 %), SPEC : 89 % (88 % à 90 %), VPP : 89,5 %, VPN : 42,8 %, stat C : 0,66 (0,66 à 0,67), kappa : 0,26 (0,24 à 0,27)
		≥ 1 demande de remboursement d'un psychiatre ou d'un neurologue	SENS : 14 % (13 % à 15 %), SPEC : 97 % (96 % à 97 %), VPP : 89,6 %, VPN : 34,8 %, stat C : 0,55 (0,55 à 0,56), kappa : 0,07 (0,07 à 0,08)

Suite à la page suivante

TABLEAU 3 (suite)
Exactitude diagnostique des études incluses

Premier auteur, année Pays	Test de référence	Algorithme(s) des bases de données administratives	Mesures de l'exactitude diagnostique (IC à 95 %) ^a
Brooks, 2021 ²³ Canada	Examen du dossier médical – diagnostic de TSA :	≥ 1 congé de l'hôpital ou ≥ 1 visite à l'urgence ou ≥ 1 chirurgie ambulatoire ou ≥ 1 demande de remboursement d'un médecin	SENS : 75,9 % (68,0 % à 83,8 %), SPEC : 98,9 % (98,6 % à 99,1 %), VPP : 42,9 % (36,0 % à 49,8 %), VPN : 99,7 % (99,6 % à 99,8 %)
	Examen manuel du dossier médical électronique à la recherche d'un diagnostic de TSA. Les cas ont été répertoriés par un infirmier qualifié responsable de l'abstraction des données des dossiers et confirmés par un médecin de famille.	≥ 1 demande de remboursement d'un médecin	SENS : 74,1 % (66,0 % à 82,2 %), SPEC : 98,9 % (98,7 % à 99,1 %), VPP : 42,6 % (35,6 % à 49,5 %), VPN : 99,7 % (99,6 % à 99,8 %)
		≥ 1 congé de l'hôpital ou ≥ 1 visite à l'urgence ou ≥ 1 chirurgie ambulatoire ou ≥ 1 demande de remboursement d'un médecin, peu importe la spécialité	SENS : 67,9 % (59,2 % à 76,5 %), SPEC : 99,0 % (98,9 % à 99,2 %), VPP : 44,7 % (37,2 % à 52,2 %), VPN : 99,6 % (99,5 % à 99,8 %)
		≥ 1 demande de remboursement d'un médecin, peu importe la spécialité	SENS : 66,1 % (57,3 % à 74,8 %), SPEC : 99,1 % (98,9 % à 99,2 %), VPP : 44,3 % (36,8 % à 51,8 %), VPN : 99,6 % (99,5 % à 99,7 %)
		≥ 1 congé de l'hôpital ou ≥ 1 visite à l'urgence ou ≥ 1 chirurgie ambulatoire ou ≥ 2 demandes de remboursement d'un médecin en 3 ans	SENS : 59,8 % (50,7 % à 68,9 %), SPEC : 99,3 % (99,1 % à 99,5 %), VPP : 49,3 % (40,9 % à 57,7 %), VPN : 99,5 % (99,4 % à 99,7 %)
		≥ 1 congé de l'hôpital ou ≥ 1 visite à l'urgence ou ≥ 1 chirurgie ambulatoire ou ≥ 2 demandes de remboursement d'un médecin en 2 ans	SENS : 57,1 % (48,0 % à 66,3 %), SPEC : 99,3 % (99,1 % à 99,5 %), VPP : 48,5 % (40,0 % à 57,0 %), VPN : 99,5 % (99,4 % à 99,7 %)
		≥ 1 congé de l'hôpital ou ≥ 1 visite à l'urgence ou ≥ 1 chirurgie ambulatoire ou ≥ 2 demandes de remboursement d'un médecin en 3 ans avec ≥ 1 demande de remboursement d'un spécialiste	SENS : 53,6 % (44,3 % à 62,8 %), SPEC : 99,4 % (99,2 % à 99,5 %), VPP : 48,4 % (39,6 % à 57,2 %), VPN : 99,5 % (99,3 % à 99,6 %)
		≥ 1 congé de l'hôpital ou ≥ 1 visite à l'urgence ou ≥ 1 chirurgie ambulatoire ou ≥ 2 demandes de remboursement d'un médecin en 2 ans avec ≥ 1 demande de remboursement d'un spécialiste	SENS : 52,7 % (43,4 % à 61,9 %), SPEC : 99,4 % (99,2 % à 99,5 %), VPP : 48,4 % (39,5 % à 57,2 %), VPN : 99,5 % (99,3 % à 99,6 %)
		≥ 1 congé de l'hôpital ou ≥ 1 visite à l'urgence ou ≥ 1 chirurgie ambulatoire ou ≥ 3 demandes de remboursement d'un médecin en 3 ans	SENS : 50,0 % (40,7 % à 59,3 %), SPEC : 99,6 % (99,4 % à 99,7 %), VPP : 56,6 % (46,8 % à 66,3 %), VPN : 99,4 % (99,3 % à 99,6 %)
		≥ 1 congé de l'hôpital ou ≥ 1 visite à l'urgence ou ≥ 1 chirurgie ambulatoire ou ≥ 3 demandes de remboursement d'un médecin en 2 ans avec ≥ 1 demande de remboursement d'un spécialiste	SENS : 49,1 % (39,8 % à 58,4 %), SPEC : 99,6 % (99,5 % à 99,7 %), VPP : 57,9 % (48,0 % à 67,8 %), VPN : 99,4 % (99,3 % à 99,6 %)
Brooks, 2021 ^{24,b} Canada	Examen du dossier médical – diagnostic de TSA :	≥ 1 congé de l'hôpital ou ≥ 1 visite à l'urgence ou ≥ 1 chirurgie ambulatoire ou ≥ 3 demandes de remboursement d'un médecin en 2 ans	SENS : 45,5 % (36,3 % à 54,8 %), SPEC : 99,6 % (99,4 % à 99,7 %), VPP : 54,3 % (44,2 % à 64,3 %), VPN : 99,4 % (99,2 % à 99,5 %)
	Examen manuel du dossier médical électronique à la recherche d'un diagnostic de TSA. Les cas ont été répertoriés par un infirmier qualifié responsable de l'abstraction des données des dossiers et confirmés par un médecin de famille.	≥ 2 demandes de remboursement d'un médecin (299 ou 315) en 3 ans	SENS : 45,5 % (36,3 % à 54,8 %), SPEC : 99,6 % (99,5 % à 99,7 %), VPP : 56,0 % (45,8 % à 66,2 %), VPN : 99,4 % (99,2 % à 99,5 %)
		≥ 2 demandes de remboursement d'un médecin (299 ou 315) en 2 ans	SENS : 33,0 % (24,4 % à 42,6 %), SPEC : 98,8 % (98,5 % à 99,0 %), VPP : 23,4 % (17,1 % à 30,8 %), VPN : 99,2 % (99,0 % à 99,4 %)
			SENS : 14,3 % (8,4 % à 22,2 %), SPEC : 99,8 % (99,7 % à 99,9 %), VPP : 44,4 % (27,9 % à 61,9 %), VPN : 99,0 % (98,8 % à 99,2 %)
			SENS : 13,4 % (7,7 % à 21,1 %), SPEC : 99,8 % (99,7 % à 99,9 %), VPP : 45,5 % (28,1 % à 63,6 %), VPN : 99,0 % (98,8 % à 99,2 %)

Suite à la page suivante

TABLEAU 3 (suite)
Exactitude diagnostique des études incluses

Premier auteur, année Pays	Test de référence	Algorithme(s) des bases de données administratives	Mesures de l'exactitude diagnostique (IC à 95 %) ^a
Burke, 2014 ²⁵ États-Unis	Examen du dossier médical à critères de classification clinique, diagnostic de TSA : Critères utilisés pour confirmer le TSA : (1) niveau 1 = descriptions comportementales fortement évocatrices d'un TSA et conformes aux critères du DSM-IV-TR ou (2) niveau 2 = diagnostic documenté par le fournisseur ou certains signes de comportements liés au TSA conformes aux critères du DSM-IV-TR (mais la description n'est pas suffisante pour la classification au niveau 1).	≥ 2 demandes de remboursement d'un médecin (299 ou 315) en 1 an	SENS : 11,6 % (6,3 % à 19,0 %), SPEC : 99,9 % (99,8 % à 99,9 %), VPP : 52,0 % (31,3 % à 72,2 %), VPN : 99,0 % (98,8 % à 99,2 %)
		≥ 3 demandes de remboursement d'un médecin (299 ou 315) en 3 ans	SENS : 2,7 % (0,6 % à 7,6 %), SPEC : 99,9 % (99,9 % à 100 %), VPP : 33,3 % (7,5 % à 70,1 %), VPN : 98,9 % (98,7 % à 99,1 %)
		≥ 3 demandes de remboursement d'un médecin (299 ou 315) en 2 ans	SENS : 2,7 % (0,6 % à 7,6 %), SPEC : 99,9 % (99,9 % à 100 %), VPP : 33,3 % (7,5 % à 70,1 %), VPN : 98,9 % (98,7 % à 99,1 %)
		≥ 3 demandes de remboursement d'un médecin (299 ou 315) en 1 an	SENS : 1,8 % (0,2 % à 6,3 %), SPEC : 100 % (99,9 % à 100 %), VPP : 33,3 % (4,3 % à 77,7 %), VPN : 98,9 % (98,7 % à 99,1 %)
		≥ 1 demande de remboursement d'un médecin (299 seulement)	SENS : 28,6 % (20,4 % à 37,9 %), SPEC : 99,9 % (99,9 % à 100 %), VPP : 86,5 % (71,2 % à 95,5 %), VPN : 99,2 % (99,0 % à 99,4 %)
		≥ 2 demandes de remboursement d'un médecin (299 seulement) en 3 ans	SENS : 12,5 % (7,0 % à 20,1 %), SPEC : 100 % (99,9 % à 100 %), VPP : 93,3 % (68,1 % à 99,8 %), VPN : 99,0 % (98,8 % à 99,2 %)
		≥ 2 demandes de remboursement d'un médecin (299 seulement) en 2 ans	SENS : 11,6 % (6,3 % à 19,0 %), SPEC : 100 % (99,9 % à 100 %), VPP : 92,9 % (66,1 % à 99,8 %), VPN : 99,0 % (98,8 % à 99,2 %)
		≥ 2 demandes de remboursement d'un médecin (299 seulement) en 1 an	SENS : 9,8 % (5,0 % à 16,9 %), SPEC : 100 % (99,9 % à 100 %), VPP : 91,7 % (61,5 % à 99,8 %), VPN : 99,0 % (98,8 % à 99,2 %)
		≥ 3 demandes de remboursement d'un médecin (299 seulement) en 3 ans	SENS : 1,8 % (0,2 % à 6,3 %), SPEC : 100 % (99,9 % à 100 %), VPP : 66,7 % (9,4 % à 99,2 %), VPN : 98,9 % (98,7 % à 99,1 %)
		≥ 3 demandes de remboursement d'un médecin (299 seulement) en 2 ans	SENS : 1,8 % (0,2 % à 6,3 %), SPEC : 100 % (99,9 % à 100 %), VPP : 66,7 % (9,4 % à 99,2 %), VPN : 98,9 % (98,7 % à 99,1 %)
		≥ 3 demandes de remboursement d'un médecin (299 seulement) en 1 an	SENS : 0,9 % (0,0 % à 4,9 %), SPEC : 100 % (100 % à 100 %), VPP : 100 % (2,5 % à 100 %), VPN : 98,9 % (98,7 % à 99,1 %)
		≥ 1 trouble lié au TSA (aucune réclamation d'assurance pour un TSA)	VPN (niveau 1 ou 2) : > 98 %
		≥ 1 réclamation d'assurance pour un TSA	VPP (niveau 1) : 43,3 % (38,2 % à 48,5 %) VPP (niveau 1 ou 2) : 74,2 % (69,4 % à 78,6 %)
		≥ 2 réclamations d'assurance pour un TSA	VPP (niveau 1) : 60,9 % (53,5 % à 68,1 %) VPP (niveau 1 ou 2) : 87,4 % (81,6 % à 91,8 %)

Suite à la page suivante

TABLEAU 3 (suite)
Exactitude diagnostique des études incluses

Premier auteur, année Pays	Test de référence	Algorithme(s) des bases de données administratives	Mesures de l'exactitude diagnostique (IC à 95 %) ^a
Coleman, 2015 ²⁶ États-Unis	Examen du dossier médical à critères de classification clinique, diagnostic de TSA : Critères utilisés pour confirmer le TSA : (1) confirmé = évaluation complète et documentée fondée sur les critères du DSM-IV; (2) probable = diagnostic posé par une source crédible, utilisation documentée du DSM-IV pour poser le diagnostic et certains comportements documentés du patient conformes aux critères du DSM-IV; ou (3) possible = rapports de seconde main d'une évaluation par un professionnel pour un TSA ou certains comportements documentés liés au TSA.	1 réclamation d'assurance ou diagnostic en consultation externe ≥ 2 réclamations d'assurance ou diagnostics en consultation externe, à au moins une journée d'intervalle	VPP ^c (confirmé) : 27 % VPP ^c (confirmé, probable et possible) : 72 % VPP ^c (confirmé) : 36 % VPP ^c (confirmé, probable et possible) : 87 %
Coo, 2017 ²⁷ Canada	Examen du dossier médical à diagnostic de TSA : Examen du dossier à la recherche d'un diagnostic de TSA par l'un de quatre fournisseurs de services en matière de comportement ou de handicap pour les enfants et les jeunes. Diagnostic déclaré par le parent : Dans le cas des enfants qui ont reçu un diagnostic administratif de TSA, mais dont aucun diagnostic confirmé ne figure au dossier, les diagnostics déclarés par les parents ont également été considérés comme de vrais positifs.	De 2 à 5 ans : ≥ 1 demande de remboursement d'un médecin ou ≥ 1 code d'éducation « TSA » ou ≥ 1 congé de l'hôpital ou ≥ 1 diagnostic dans un centre de traitement pour adolescents De 6 à 9 ans : ≥ 1 demande de remboursement d'un médecin ou ≥ 1 code d'éducation « TSA » ≥ 2 demandes de remboursement d'un médecin ou ≥ 1 code d'éducation « TSA » ≥ 2 demandes de remboursement d'un médecin ou ≥ 1 code d'éducation « TSA » ≥ 2 demandes de remboursement d'un médecin ou ≥ 2 codes d'éducation « TSA » ≥ 2 demandes de remboursement d'un médecin ou ≥ 2 codes d'éducation « TSA » ≥ 1 demande de remboursement d'un médecin ou ≥ 1 code d'éducation « TSA » ≥ 1 demande de remboursement d'un médecin ou ≥ 2 codes d'éducation « TSA » ≥ 2 demandes de remboursement d'un médecin ou ≥ 2 codes d'éducation « TSA » ≥ 1 demande de remboursement d'un médecin ou ≥ 1 code d'éducation « TSA » ≥ 2 codes d'éducation « TSA »	De 2 à 5 ans : SENS : 88 % (84 % à 91 %), VPP ^d minimum : 73 % (68 % à 77 %) De 6 à 9 ans : SENS : 88 % (83 % à 91 %), VPP ^d minimum : 73 % (69 % à 78 %) SENS : 85 % (80 % à 88 %), VPP ^d minimum : 73 % (68 % à 77 %) SENS : 57 % (51 % à 62 %), VPP ^d minimum : 89 % (84 % à 93 %) SENS : 50 % (45 % à 56 %), VPP ^d minimum : 89 % (83 % à 93 %) De 6 à 9 ans : SENS : 90 % (88 % à 93 %), VPP ^d minimum : 65 % (61 % à 68 %) De 6 à 9 ans : SENS : 89 % (86 % à 92 %), VPP ^d minimum : 65 % (61 % à 68 %) SENS : 88 % (85 % à 90 %), VPP ^d minimum : 65 % (62 % à 69 %) SENS : 84 % (81 % à 87 %), VPP ^d minimum : 78 % (75 % à 81 %) SENS : 81 % (78 % à 84 %), VPP ^d minimum : 80 % (77 % à 84 %) SENS : 77 % (73 % à 80 %), VPP ^d minimum : 66 % (62 % à 69 %) SENS : 68 % (64 % à 72 %), VPP ^d minimum : 87 % (84 % à 90 %) SENS : 66 % (62 % à 70 %), VPP ^d minimum : 88 % (85 % à 91 %)

Suite à la page suivante

TABLEAU 3 (suite)
Exactitude diagnostique des études incluses

Premier auteur, année Pays	Test de référence	Algorithme(s) des bases de données administratives	Mesures de l'exactitude diagnostique (IC à 95 %) ^a
Dodds, 2009 ²⁸ Canada	Diagnostic clinique : Diagnostic clinique posé par une équipe de spécialistes du TSA, à l'aide de l'ADI-R (Autism Diagnostic Interview à Revised), de l'ADOS (Autism Diagnostic Observation Schedule) et du jugement clinique fondé sur le DSM-IV-TR.	≥ 2 demandes de remboursement d'un médecin	SENS : 58 % (54 % à 62 %), VPP ^d minimum : 83 % (79 % à 86 %)
		De 10 à 14 ans :	De 10 à 14 ans :
		≥ 1 demande de remboursement d'un médecin ou ≥ 1 code d'éducation « TSA » ou ≥ 1 congé de l'hôpital ou ≥ 1 diagnostic dans un centre de traitement pour adolescents	SENS : 88 % (85 % à 90 %), VPP ^d minimum : 60 % (57 % à 63 %)
		≥ 1 demande de remboursement d'un médecin ou ≥ 1 code d'éducation « TSA »	SENS : 86 % (83 % à 88 %), VPP ^d minimum : 61 % (58 % à 63 %)
		≥ 1 demande de remboursement d'un médecin ou ≥ 2 codes d'éducation « TSA »	SENS : 84 % (82 % à 87 %), VPP ^d minimum : 62 % (59 % à 64 %)
		≥ 2 demandes de remboursement d'un médecin ou ≥ 1 code d'éducation « TSA »	SENS : 80 % (77 % à 83 %), VPP ^d minimum : 70 % (67 % à 73 %)
		≥ 2 demandes de remboursement d'un médecin ou ≥ 2 codes d'éducation « TSA »	SENS : 78 % (75 % à 81 %), VPP ^d minimum : 72 % (69 % à 75 %)
		≥ 1 demande de remboursement d'un médecin	SENS : 73 % (69 % à 76 %), VPP ^d minimum : 64 % (60 % à 70 %)
		≥ 1 code d'éducation « TSA »	SENS : 73 % (70 % à 76 %), VPP ^d minimum : 75 % (72 % à 78 %)
		≥ 2 codes d'éducation « TSA »	SENS : 69 % (66 % à 72 %), VPP ^d minimum : 78 % (74 % à 81 %)
		≥ 2 demandes de remboursement d'un médecin	SENS : 56 % (52 % à 59 %), VPP ^d minimum : 78 % (75 % à 82 %)
Hagberg, 2017 ²⁹ Royaume-Uni	Examen du dossier médical = diagnostic de TSA : Examen du dossier médical original pour confirmer le diagnostic, ce qui comprenait des lettres cliniques détaillées de l'hôpital, des rapports de consultation, des évaluations de la parole et du langage, et des rapports de spécialistes.	≥ 1 congé de l'hôpital ou ≥ 1 demande de remboursement d'un médecin ou ≥ 1 diagnostic en consultation externe de santé mentale	SENS : 69,3 %, SPEC : 77,3 %, stat C : 0,76
		≥ 1 congé de l'hôpital ou ≥ 1 demande de remboursement d'un médecin	SENS : 62,5 %, SPEC : 83,0 %, stat C : 0,74
		≥ 1 demande de remboursement d'un médecin	SENS : 59,7 %, SPEC : 85,2 %, stat C : 0,72
		≥ 1 congé de l'hôpital ou ≥ 2 demandes de remboursement d'un médecin ou ≥ 2 diagnostics en consultation externe de santé mentale	SENS : 42,6 %, SPEC : 88,6 %, stat C : 0,67
		≥ 1 congé de l'hôpital ou ≥ 2 demandes de remboursement d'un médecin	SENS : 36,9 %, SPEC : 93,2 %, stat C : 0,65
		≥ 1 diagnostic en consultation externe de santé mentale	SENS : 16,5 %, SPEC : 92,0 %, stat C : 0,54
		≥ 1 diagnostic au congé de l'hôpital	SENS : 11,9 %, SPEC : 97,7 %, stat C : 0,55

Suite à la page suivante

TABLEAU 3 (suite)
Exactitude diagnostique des études incluses

Premier auteur, année Pays	Test de référence	Algorithme(s) des bases de données administratives	Mesures de l'exactitude diagnostique (IC à 95 %) ^a
Lauritsen, 2010 ³⁰ Danemark	Examen du dossier médical = critères de classification clinique : Version modifiée du guide de codification des CDC fondée sur le DSM-IV, c'est-à-dire présence d'au moins un critère social et un critère de communication ou de comportement, sans diagnostic, antécédents ou descriptions comportementales contredisant la présence d'un TSA.	1 diagnostic en milieu hospitalier ou en consultation externe de psychiatrie	VPP : 97 % (96 % à 99 %)
Surén, 2019 ³¹ Norvège	Examen du dossier médical = critères de classification clinique : Examen des dossiers du patient pour confirmer que l'enfant satisfait aux critères de diagnostic de la CIM-10; comprend les résultats des entrevues et des tests normalisés ainsi que les diagnostics reçus.	≥ 1 diagnostic	VPP : 86 % (83 % à 89 %)
TDAH			
Daley, 2014 ³² États-Unis	Examen du dossier médical = critères de classification clinique, diagnostic de TDAH, liste de contrôle normalisée pour le dépistage : Critères utilisés pour confirmer le TDAH : (1) définition 1 = diagnostic d'un clinicien pendant la période de référence ou de suivi; (2) définition 2 = diagnostic d'un clinicien pendant la période de référence ou de suivi, cas prévalents exclus; (3) définition 3 = diagnostic d'un clinicien, cas prévalents exclus, résultats positifs à au moins une liste de contrôle pour le dépistage du TDAH; (4) définition 4 = diagnostic d'un clinicien, cas prévalents exclus, présence d'au moins 6 des 9 symptômes d'inattention et/ou 6 des 9 symptômes d'hyperactivité ou d'impulsivité du DSM-IV; (5) définition 5 = diagnostic d'un clinicien, cas prévalents exclus, résultats positifs à au moins une liste de contrôle pour le dépistage ou présence d'au moins 6 des 9 symptômes d'inattention et/ou 6 des 9 symptômes d'hyperactivité ou d'impulsivité du DSM-IV.	2 diagnostics en consultation externe, à un intervalle de 7 à 365 jours (cas incident de TDAH)	Enfants de 3 à 5 ans au moment du diagnostic : VPP ^e (définition 1) : 89,8 % (80,6 % à 99,0 %) VPP ^e (définition 2) : 71,5 % (56,5 % à 86,4 %) VPP ^e (définition 3) : 48,9 % (33,4 % à 64,3 %) VPP ^e (définition 4) : 32,8 % (17,1 % à 48,5 %) VPP ^e (définition 5) : 65,8 % (52,2 % à 79,4 %) Enfants de 6 à 9 ans au moment du diagnostic : VPP ^e (définition 1) : 94,2 % (89,8 % à 98,5 %) VPP ^e (définition 2) : 73,6 % (65,6 % à 81,6 %) VPP ^e (définition 3) : 59,1 % (50,8 % à 67,5 %) VPP ^e (définition 4) : 30,9 % (22,2 % à 39,6 %) VPP ^e (définition 5) : 68,5 % (60,8 % à 76,1 %)

Suite à la page suivante

TABLEAU 3 (suite)
Exactitude diagnostique des études incluses

Premier auteur, année Pays	Test de référence	Algorithme(s) des bases de données administratives	Mesures de l'exactitude diagnostique (IC à 95 %) ^a
Gruschow, 2016 ³³ États-Unis	Définition de cas clinique ^b : Patients avec TDAH confirmé si : ≥ 3 visites liées au TDAH, ou 1 ou 2 visites liées au TDAH ou un diagnostic figurant dans la liste des problèmes et un médicament prescrit pour le TDAH, ou données d'une source indépendante, répertoriées par un examen manuel du dossier de santé électronique, qui confirment le statut du cas de TDAH. Patients sans TDAH confirmé lorsque des données provenant d'une source indépendante, répertoriées par un examen manuel du dossier de santé électronique, indiquent que le patient n'a pas de TDAH.	≥ 1 diagnostic en milieu hospitalier ou en consultation externe, ou diagnostic figurant dans la liste des problèmes ^c	SENS : 96 % à 97 % (95 % à 97 %), SPEC : 98 % à 99 % (97 % à 99 %), VPP : 83 % à 98 % (81 % à 99 %), VPN : 99 % (99 % à 99 %) ^d , Kappa : 0,87 (0,75 à 0,99)
Mohr-Jensen, 2016 ³³ Danemark	Examen du dossier médical = critères de classification clinique : Les dossiers des patients ont été systématiquement notés en fonction de la présence des critères de la CIM-10 pour le trouble hyperkinétique, et les cas étaient confirmés lorsque les patients présentaient ≥ 6 symptômes d'inattention, ≥ 3 symptômes d'hyperactivité et ≥ 1 symptôme d'impulsivité.	1 diagnostic en milieu hospitalier ou en consultation externe de psychiatrie	VPP : 86,8 %
Morkem, 2020 ³⁵ Canada	Examen du dossier médical = diagnostic de TDAH : Examen du dossier médical électronique à la recherche d'un diagnostic de TDAH.	≥ 1 visite médicale (code CIM) et ≥ 1 ordonnance de médicament lié au TDAH ou ≥ 2 visites médicales (code CIM)	VPP : 95,9 % (92,6 % à 98,0 %), VPN : 96,3 % (93,2 % à 98,3 %)

Abréviations : CDC, Centers for Disease Control and Prevention; CIM, Classification internationale des maladies; CIM-10, Classification internationale des maladies, dixième révision; DSM, *Manuel diagnostique et statistique des troubles mentaux*; DSM-IV, *Manuel diagnostique et statistique des troubles mentaux*, quatrième révision; DSM-IV-TR, *Manuel diagnostique et statistique des troubles mentaux*, quatrième révision, texte révisé; IC, intervalle de confiance; SENS, sensibilité; SPEC, spécificité; stat C, statistique C; TDAH, trouble du déficit de l'attention/hyperactivité; TSA, trouble du spectre de l'autisme; VPN, valeur prédictive négative; VPP, valeur prédictive positive.

^a Le cas échéant.

^b Cette étude a également mis à l'essai des algorithmes qui comprenaient des renseignements d'identification des cas tirés d'une recherche par mot-clé du profil cumulatif des patients dans le dossier de santé électronique; toutefois, ces algorithmes n'ont pas été inclus, car ils ne répondraient pas à la définition d'un algorithme de base de données administratives utilisée dans le cadre de la revue.

^c Les cas où l'information était insuffisante ont été exclus des calculs de la VPP.

^d Les participants dont le statut de cas n'a pas pu être déterminé ont été considérés comme de faux positifs aux fins de ce calcul.

^e Tous les cas échantillonnes ont été pondérés par l'inverse de leur probabilité de sélection.

^f La définition de cas clinique reposait sur un examen du dossier médical ou du dossier de santé électronique et n'était pas aussi rigoureuse que les critères de classification clinique.

^g Liste des problèmes actuels ou passés dans le dossier de santé électronique du patient.

^h La fourchette dépend des hypothèses pour les cas non concluants, c'est-à-dire les vrais cas de TDAH par rapport aux vrais cas sans TDAH.

9 ans, le fait d'exiger au moins un code provenant des demandes de remboursement des médecins ou des données sur l'éducation plutôt qu'au moins un code provenant des demandes de remboursement des médecins a entraîné une augmentation substantielle de la sensibilité (de 77 % à 89 %) et une légère diminution de la VPP (de 66 % à 65 %). Une tendance semblable a été observée chez les enfants plus âgés, soit le groupe des 10 à 14 ans.

Études sur le TDAH

Dans le cas des études sur le TDAH, l'exactitude diagnostique des algorithmes mis à l'essai a été résumée de la même façon; toutefois, aucune des études sur le TDAH n'a utilisé de données sur l'éducation en plus des données administratives sur la santé. Par conséquent, il n'a pas été possible de déterminer si ces données présentent des avantages.

Selon l'algorithme de base de données administratives sur la santé

Les quatre études ont mis à l'essai un seul algorithme chacune et en ont présenté les résultats; tous les algorithmes comprenaient des codes de diagnostic provenant d'une seule source de données administratives³²⁻³⁵. De ce fait, on n'a pas disposé de suffisamment d'information pour déterminer si le fait d'exiger plus ou moins de codes de diagnostic ou d'utiliser des sources de données supplémentaires pour repérer les cas de TDAH avait une incidence.

Selon le test de référence

Une des quatre études faisait varier les critères diagnostiques requis pour le test de référence et présentait les résultats en fonction d'exigences dont le degré de sévérité variait pour confirmer les cas incidents de TDAH³². Le fait d'exiger davantage d'éléments d'information au dossier à l'appui du cas incident de TDAH a eu pour effet de faire passer la VPP de 71,5 % à 32,8 % chez les enfants âgés de 3 à 5 ans au moment du diagnostic et de 73,6 % à 30,9 % chez les enfants de 6 à 9 ans.

Analyse

Au total, quatorze études ont répondu à nos critères d'admissibilité²²⁻³⁵; dix portant sur la validation d'algorithmes de bases de données administratives pour repérer les cas de TSA²²⁻³¹ et quatre, pour les cas de TDAH³²⁻³⁵. Six des quatorze études ont été menées au Canada avec, de façon

générale, une qualité des rapports de ces six études supérieure à celle des études d'autres pays et un risque de biais inférieur^{22-24,27,28,35}. Aucune étude sur le TSAF ne répondait aux critères d'admissibilité de notre revue. Parmi les autres lacunes importantes relevées, mentionnons l'absence d'études de validation chez les adultes et l'identification des cas incidents plutôt que prévalents.

Bien que l'on ait tenté d'utiliser les données administratives sur la santé pour estimer la prévalence du TSAF au Canada³⁶, le travail a été effectué en l'absence d'algorithmes de base de données administratives validés. L'absence d'études de validation publiées sur le TSAF pourrait être attribuable à plusieurs problèmes fondamentaux liés à l'établissement d'un diagnostic de TSAF, notamment :

- la nécessité d'une évaluation multidisciplinaire et de la connaissance de l'exposition prénatale à l'alcool^{37,38};
- l'absence de critères diagnostiques ou de description détaillée pour le diagnostic de « trouble neurodéveloppemental associé à l'exposition prénatale à l'alcool » dans le *Manuel diagnostique et statistique des troubles mentaux*, cinquième édition^{1,39};
- la nature non spécifique des codes de diagnostic de la Classification internationale des maladies qui peuvent être utilisés pour un diagnostic principal de TSAF³⁹ et la spécificité des codes requis pour saisir les entités diagnostiques associées au TSAF, qui n'est pas toujours prise en charge par les bases de données administratives sur la santé.

Vu la nature évolutive des pratiques de diagnostic et de la codification se rapportant au TSAF, l'utilisation d'algorithmes de bases de données administratives pour repérer les cas de TSAF pose des difficultés particulières à l'heure actuelle.

Les résultats des évaluations de la qualité ont révélé un risque de biais « élevé » ou « incertain » dans au moins un domaine de l'outil QUADAS-2 pour 12 des 14 études^{22,25-35}, ce qui signifie que les mesures de l'exactitude diagnostique doivent être interprétées avec prudence. En outre, en raison de la grande hétérogénéité des études retenues dans la façon dont elles ont été conçues et menées, il n'a pas été possible de faire une synthèse quantitative des résultats sur l'exactitude

diagnostique des algorithmes mis à l'essai. Il convient de souligner plus particulièrement les différences dans la méthode de sélection initiale des cas (selon les codes de diagnostic dans la base de données administrative ou selon le test de référence) et l'inclusion ou l'exclusion des sujets ne présentant pas le trouble étudié. Pour garantir l'impartialité des estimations de l'exactitude diagnostique, il faut que la prévalence de la maladie dans la cohorte de validation se rapproche de la prévalence dans la population⁴⁰. Il en est ainsi lorsqu'un test de référence approprié est appliqué (soit un test de référence qui classe correctement les sujets qui présentent l'affection étudiée et les sujets qui ne la présentent pas) et que les patients sont échantillonés aléatoirement, idéalement dans la population générale. Par ailleurs, pour calculer les principales estimations de l'exactitude diagnostique nécessaires à l'évaluation des caractéristiques d'un test diagnostique, il faut des sujets qui présentent l'affection étudiée et des sujets qui ne la présentent pas afin de remplir les quatre cellules d'un tableau de contingence.

Malheureusement, la plupart des études incluses dans notre revue (huit sur quatorze) ont utilisé des codes de diagnostic dans la base de données administratives pour sélectionner au départ les cas présentant le trouble étudié seulement^{26,27,29-32,34,35}. Cette façon de faire peut produire des estimations biaisées de l'exactitude diagnostique, étant donné que la prévalence sous-jacente des troubles étudiés est inconnue, en plus de limiter les mesures de l'exactitude diagnostique qu'il est possible de calculer à la VPP uniquement⁴⁰. Bien que la VPP indique la probabilité de résultats de test faussement positifs, elle ne fournit pas à elle seule d'information sur la probabilité de résultats de test faussement négatifs ou sur le nombre de cas manqués par l'algorithme.

De plus, même si sept des quatorze études incluaient des sujets ne présentant pas le trouble étudié^{22-25,28,33,35}, deux ont procédé à leur échantillonnage dans des cliniques spécialisées ou auprès de fournisseurs de services^{22,28} et deux autres ont suréchantilloné des enfants plus susceptibles d'avoir le trouble étudié^{25,33}, ces deux méthodes d'échantillonnage pouvant générer une VPP faussement élevée en raison de la prévalence élevée des cas. En outre, seulement quatre de ces études ont fait état des quatre mesures clés de l'exactitude

diagnostique calculables au moyen de cette approche^{22-24,33}.

Malgré les limites susmentionnées, les résultats de notre revue donnent à penser que l'augmentation du nombre de codes de diagnostic de TSA requis dans la base de données sur les demandes de remboursement des médecins augmente la spécificité et la VPP d'un algorithme, au détriment de la sensibilité. De plus, l'utilisation de plusieurs sources de données administratives sur la santé dans un algorithme conçu pour recenser les cas de TSA (hôpitaux, demandes de remboursement des médecins et services de santé mentale) pourrait accroître la sensibilité sans trop compromettre la spécificité et la VPP, la base de données sur les demandes de remboursement des médecins étant la meilleure source unique.

Par ailleurs, les résultats d'une étude ont montré que l'ajout de données sur l'éducation, en combinaison avec les données tirées des demandes de remboursement des médecins, pourrait améliorer l'identification des cas (sensibilité) chez les enfants et les jeunes d'âge scolaire sans trop compromettre la précision (VPP)²⁷. Toutefois, l'absence de sujets sans TSA dans cette étude limitait les mesures de l'exactitude diagnostique calculables. Par conséquent, d'autres études sont nécessaires pour évaluer l'incidence complète de l'inclusion de données sur l'éducation en combinaison avec les données tirées des demandes de remboursement des médecins dans les algorithmes des bases de données administratives pour déterminer les cas de TSA.

En raison de la nature des études sur le TDAH retenues, l'information fournie n'était pas suffisante pour évaluer l'incidence du nombre de codes de diagnostic ou de sources de données supplémentaires sur l'exactitude diagnostique des algorithmes. Toutefois, selon les mesures du rendement déclarées, certaines données semblent indiquer que les cas de TDAH pourraient être recensés au moyen de sources de données administratives sur la santé.

Pour combler les lacunes révélées par cette revue et remédier aux problèmes observés dans la qualité des rapports et le risque de biais, il faudra mener d'autres études de grande qualité validant l'utilisation d'algorithmes de bases de données administratives pour repérer les cas des

troubles neurodéveloppementaux sélectionnés. Ces problèmes ne sont pas propres à ce domaine d'étude, et des lignes directrices sur la façon de mener de telles études de validation et d'en communiquer les résultats ont déjà été publiées^{12,20,40}. À la lumière de ce qui précède, nous recommandons aux auteurs qui souhaitent valider les algorithmes des bases de données administratives utilisés pour identifier les cas de suivre les recommandations publiées sur les méthodes d'étude^{20,40} ainsi que les lignes directrices sur la production de rapports¹².

Le fait que les limites entre ces troubles sont souvent floues et que la présence de troubles concomitants est fréquente constitue une autre difficulté associée à l'utilisation de codes de diagnostic précis pour les troubles neurodéveloppementaux comme le TSA, le TDAH et le TSAF. Alors que, dans le passé, les troubles neurodéveloppementaux étaient diagnostiqués de manière catégorique sur la base d'un ensemble de signes et de symptômes, on observe un nombre croissant de publications soulignant la nécessité d'adopter de nouvelles approches diagnostiques qui conceptualisent ces troubles comme appartenant à un continuum neurodéveloppemental⁴¹. Ce changement aura des répercussions importantes sur la classification de ces troubles, la pratique clinique, la recherche et la surveillance.

Points forts et limites

Les points forts de cette revue sont les suivants :

- l'enregistrement prospectif de la revue dans le registre prospectif international des revues systématiques (PROSPERO), qui contribue à réduire le risque de biais dans la conduite et la rédaction des rapports des revues systématiques⁴²;
- l'élaboration, par une bibliothécaire de références chevronnée, d'une stratégie de recherche de la littérature comprenant une recherche systématique dans plusieurs bases de données, dans la littérature grise et dans les listes de référence des articles inclus;
- une évaluation rigoureuse de la qualité des rapports ainsi que du risque de biais et de l'applicabilité de chaque étude retenue à l'aide de la liste de vérification modifiée STARD¹² et de l'outil QUADAS-2²⁰ respectivement;

- l'utilisation des normes PRISMA pour assurer la transparence et l'exhaustivité du rapport de la revue¹⁷.

Il convient toutefois de souligner certaines limites, notamment :

- des difficultés liées à la recherche exhaustive d'études axées sur les algorithmes des bases de données administratives, puisqu'elles ne sont pas bien répertoriées dans les bases de données que nous avons consultées (il n'existe aucune vedette-matière médicale de type « base de données administratives »);
- le risque de biais linguistique, puisque les études publiées dans une langue autre que l'anglais ou le français n'ont pas été prises en considération, ainsi que le risque de biais de publication, puisque les études de validation ayant obtenu de piètres résultats pourraient être moins susceptibles d'être publiées;
- l'hétérogénéité importante entre les études retenues, qui a rendu impossible la réalisation d'analyses quantitatives, que ce soit une méta-régression ou une méta-analyse.

Conclusion

À notre connaissance, il s'agit de la première revue ayant systématiquement évalué et examiné les données empiriques sur la validité des algorithmes des bases de données administratives utilisés pour repérer les cas de TSA, de TDAH et de TSAF. Bien que quelques études aient validé des algorithmes pour la détermination des cas de TSA et de TDAH, aucune n'a été effectuée pour le TSAF à ce jour. La grande hétérogénéité entre les études incluses a limité notre capacité à effectuer des analyses quantitatives. De telles analyses permettraient de renforcer davantage les données probantes sur les algorithmes les plus performants aux fins de la surveillance des troubles neurodéveloppementaux et de la recherche en la matière, pourvu que les études disponibles soient de qualité suffisante.

Néanmoins, certaines données semblent indiquer que les cas de TSA et de TDAH sont identifiables à l'aide de données administratives, quoique l'information sur la capacité à faire une distinction fiable entre les personnes atteintes ou non du trouble étudié soit limitée. Compte tenu des variations observées dans la qualité des rapports et du risque de biais, d'autres

études de validation de grande qualité sont nécessaires. Afin d'optimiser l'utilité des études futures, nous recommandons aux auteurs de suivre les recommandations publiées sur la conception et la réalisation des études^{20,40} ainsi que les lignes directrices sur la production de rapports d'études de validation portant sur des données administratives¹².

Remerciements

Les auteures tiennent à remercier Katherine Merucci, bibliothécaire de référence à la Bibliothèque de la santé de la Direction générale des services de gestion de Santé Canada et de l'Agence de la santé publique

du Canada, qui a élaboré et géré les stratégies de recherche dans les bases de données et la littérature grise.

Aucun financement, que ce soit sous forme de subvention ou sous une autre forme de soutien à la recherche, n'a été obtenu pour cette étude.

Conflits d'intérêts

Aucun.

Contributions des auteures et avis

CL et SO ont conçu l'étude. SO et SP ont participé à l'élaboration de la stratégie de

recherche et à l'extraction des articles. CL, ML et SO ont procédé au tri de la littérature. ML et SO ont extrait les données et évalué la qualité des rapports, le risque de biais et l'applicabilité. Toutes les auteures ont analysé ou interprété les données. SO et SP ont rédigé le manuscrit. Toutes les auteures ont contribué à la première ébauche et aux révisions du manuscrit.

Le contenu de l'article et les points de vue qui y sont exprimés n'engagent que les auteures; ils ne correspondent pas nécessairement à ceux du gouvernement du Canada.

ANNEXE A
Stratégie de recherche électronique

Recherche INITIALE

Base(s) de données : Ovid MEDLINE(R) ALL, de 1946 au 26 août 2019

Stratégie de recherche :

Nº	Recherches	Résultats
1	exp autism spectrum disorder/ ou (autism ou autistic ou (asperger* adj (syndrome ou disorder ou disease)) ou kanner* syndrome ou childhood disintegrative disorder ou (pervasive adj2 developmental disorder*) ou heller* syndrome ou disintegrative psychosis).tw,kf,kw.	48476
2	Attention Deficit Disorder with Hyperactivity/ ou "Attention Deficit and Disruptive Behavior Disorders"/ ou ((attention deficit adj4 disorder*) ou (hyperkinetic adj2 (disorder ou syndrome)) ou minimal brain dysfunction ou adhd ou addh).tw,kf,kw.	40083
3	Fetal Alcohol Spectrum Disorders/ ou (f?etal alcohol ou (alcohol related adj3 (neurodevelopment ou birth)) ou "Neurobehavioral disorder associated with prenatal alcohol exposure" ou "Growth Retardation, Facial Abnormalities, and Central Nervous System Dysfunction").tw,kw,kf.	5699
4	ou/1-3	90368
5	diagnosis/ ou incidence/ ou prevalence/ ou (diagnos* ou incidence* ou prevalence* ou new case?).tw,kw,kf.	3507353
6	4 et 5	27896
7	exp autism spectrum disorder/di, dg, ep ou Attention Deficit Disorder with Hyperactivity/di, dg, ep ou "Attention Deficit and Disruptive Behavior Disorders"/di, dg, ep ou Fetal Alcohol Spectrum Disorders/di, dg, ep	22499
8	6 ou 7 [Troubles du développement]	39288
9	exp medical records/ ou international classification of diseases/ ou exp diagnostic techniques, neurological/ ou exp clinical laboratory techniques/ ou "diagnostic techniques and procedures"/	3030824
10	((patient* ou medical ou health) adj3 record*) ou ((diagnos* ou defin* ou classificat*) adj2 disease list) ou ((self report* ou standard*) adj2 measure*) ou (icd adj (cod* ou classification*)) ou (international adj3 classification)).tw,kw,kf.	248712
11	medicaid/ ou birth certificates/ ou death certificates/ ou hospital records/ ou insurance claim reporting/ ou exp insurance, health/ ou databases, factual/ ou information systems/ ou databases as topic/ ou database management systems/ ou software/ ou insurance claim review/ ou patient discharge/ ou exp registries/ ou utilization review/	472574
12	((administrat* ou physician* ou inpatient* ou emergency* ou hospital* ou clinic ou clinics ou pharmac* ou insurance) adj4 (admission* ou data ou dataset* ou database* ou data base? ou data bank? ou claim* ou billing* ou record* ou utilizat* ou utilisat*)) ou ((claim* ou discharg*) adj2 data*) ou ((database? ou databank? ou data base? ou data bank?) adj4 (factual ou administrat* ou claim? ou register* ou registr* ou topic? ou system?)) ou (claim? adj2 (analy* ou review? ou physician? ou pharmac* ou drug?)) ou (insurance adj2 (claim* ou audit*)) ou medicaid ou (health* adj2 plan) ou ((death* ou birth*) adj1 (certificate* ou record?)).tw,kw,kf. ou (database? ou databank? ou data base? ou data bank?).ti.	290378
13	ou/9-12 [Méthodes de diagnostic]	3827029
14	8 et 13	5662
15	exp Diagnostic Errors/ ou Diagnosis, Differential/ ou "Predictive Value of Tests"/ ou "Sensitivity and Specificity"/ ou ROC Curve/ ou Area under Curve/ ou Bayes Theorem/ ou algorithms/ ou validation studies as topic/	1272706
16	((diagnos* adj3 (schedul* ou clinical* ou technique* ou procedur* ou assess* ou standard* ou error* ou false ou incorrect* ou wrong* ou correct*) ou misdiagnos* ou (clinical* adj (standard* ou criteri* ou measure* ou classifi* ou technique* ou assess*))) ou ((positive ou negative) adj2 (predict* ou false)) ou sensitiv* ou specif* ou accura* ou valid* ou reliab* ou agree* ou concord* ou misclass* ou ((case ou cases) adj2 ascertain*) ou algorithm? ou (bayes* adj (theorem ou analysis ou approach ou forecast ou method ou prediction)) ou (roc adj (curve ou analysis)) ou receiver operating characteristic).tw,kw,kf.	5708365
17	15 ou 16	6369226
18	14 et 17	2476
19	limit 18 to (yr="1995-2019" et (english ou french))	2196

ANNEXE A (suite)
Stratégie de recherche électronique

Recherche – Première mise à jour

Base(s) de données : Ovid MEDLINE(R) ALL, de 1946 au 10 juillet 2020

Stratégie de recherche :

Nº	Recherches	Résultats
1	exp autism spectrum disorder/ ou (autism ou autistic ou (asperger* adj (syndrome ou disorder ou disease)) ou kanner* syndrome ou childhood disintegrative disorder ou (pervasive adj2 developmental disorder*) ou heller* syndrome ou disintegrative psychosis).tw,kf,kw.	53121
2	Attention Deficit Disorder with Hyperactivity/ ou "Attention Deficit and Disruptive Behavior Disorders"/ ou ((attention deficit adj4 disorder*) ou (hyperkinetic adj2 (disorder ou syndrome)) ou minimal brain dysfunction ou adhd ou addh).tw,kf,kw.	42517
3	Fetal Alcohol Spectrum Disorders/ ou (f?etal alcohol ou (alcohol related adj3 (neurodevelopment ou birth)) ou "Neurobehavioral disorder associated with prenatal alcohol exposure" ou "Growth Retardation, Facial Abnormalities, and Central Nervous System Dysfunction").tw,kw,kf.	5917
4	ou/1-3	97183
5	diagnosis/ ou incidence/ ou prevalence/ ou (diagnos* ou incidence* ou prevalence* ou new case?).tw,kw,kf.	3713126
6	4 et 5	30299
7	exp autism spectrum disorder/di, dg, ep ou Attention Deficit Disorder with Hyperactivity/di, dg, ep ou "Attention Deficit and Disruptive Behavior Disorders"/di, dg, ep ou Fetal Alcohol Spectrum Disorders/di, dg, ep	23876
8	6 ou 7 [Troubles du développement]	42246
9	exp medical records/ ou international classification of diseases/ ou exp diagnostic techniques, neurological/ ou exp clinical laboratory techniques/ ou "diagnostic techniques and procedures"/	3106548
10	((patient* ou medical ou health) adj3 record*) ou ((diagnos* ou defin* ou classificat*) adj2 disease list) ou ((self report* ou standard*) adj2 measure*) ou (icd adj (cod* ou classification*)) ou (international adj3 classification)).tw,kw,kf.	268618
11	medicaid/ ou birth certificates/ ou death certificates/ ou hospital records/ ou insurance claim reporting/ ou exp insurance, health/ ou databases, factual/ ou information systems/ ou databases as topic/ ou database management systems/ ou software/ ou insurance claim review/ ou patient discharge/ ou exp registries/ ou utilization review/	498782
12	((administrat* ou physician* ou inpatient* ou emergency* ou hospital* ou clinic ou clinics ou pharmac* ou insurance) adj4 (admission* ou data ou dataset* ou database* ou data base? ou data bank? ou claim* ou billing* ou record* ou utilizat* ou utilisat*)) ou ((claim* ou discharg*) adj2 data*) ou ((database? ou databank? ou data base? ou data bank?) adj4 (factual ou administrat* ou claim? ou register* ou registr* ou topic? ou system?)) ou (claim? adj2 (analy* ou review? ou physician? ou pharmac* ou drug?)) ou (insurance adj2 (claim* ou audit*)) ou medicaid ou (health* adj2 plan) ou ((death* ou birth*) adj1 (certificate* ou record*)).tw,kw,kf. ou (database? ou databank? ou data base? ou data bank?).ti.	312830
13	ou/9-12 [Diagnostic]	3955207
14	8 et 13	6161
15	exp Diagnostic Errors/ ou Diagnosis, Differential/ ou "Predictive Value of Tests"/ ou "Sensitivity and Specificity"/ ou ROC Curve/ ou Area under Curve/ ou Bayes Theorem/ ou algorithms/ ou validation studies as topic/	1317678
16	((diagnos* adj3 (schedul* ou clinical* ou technique* ou procedur* ou assess* ou standard* ou error* ou false ou incorrect* ou wrong* ou correct*)) ou misdiagnos* ou (clinical* adj (standard* ou criteri* ou measure* ou classifi* ou technique* ou assess*)) ou ((positive ou negative) adj2 (predict* ou false)) ou sensitiv* ou specif* ou accura* ou valid* ou reliab* ou agree* ou concord* ou misclass* ou ((case ou cases) adj2 ascertain*) ou algorithm? ou (bayes* adj (theorem ou analysis ou approach ou forecast ou method ou prediction)) ou (roc adj (curve ou analysis)) ou receiver operating characteristic).tw,kw,kf.	6045219
17	15 ou 16	6721820
18	14 et 17	2712
19	limit 18 to (yr="1995-Current" et (english ou french))	2429
20	(201908* ou 201909* ou 201910* ou 201911* ou 201912* ou 202*).ez.	1099141
21	19 et 20	99

ANNEXE A (suite)
Stratégie de recherche électronique

Recherche – Seconde mise à jour

Base(s) de données : Ovid MEDLINE(R) ALL, de 1946 au 30 mars 2021

Stratégie de recherche :

Nº	Recherches	Résultats
1	exp autism spectrum disorder/ ou (autism ou autistic ou (asperger* adj (syndrome ou disorder ou disease)) ou kanner* syndrome ou childhood disintegrative disorder ou (pervasive adj2 developmental disorder*) ou heller* syndrome ou disintegrative psychosis).tw,kf,kw.	57062
2	Attention Deficit Disorder with Hyperactivity/ ou "Attention Deficit and Disruptive Behavior Disorders"/ ou ((attention deficit adj4 disorder*) ou (hyperkinetic adj2 (disorder ou syndrome)) ou minimal brain dysfunction ou adhd ou addh).tw,kf,kw.	44499
3	Fetal Alcohol Spectrum Disorders/ ou (f?etal alcohol ou (alcohol related adj3 (neurodevelopment ou birth)) ou "Neurobehavioral disorder associated with prenatal alcohol exposure" ou "Growth Retardation, Facial Abnormalities, and Central Nervous System Dysfunction").tw,kw,kf.	6077
4	ou/1-3	102829
5	diagnosis/ ou incidence/ ou prevalence/ ou (diagnos* ou incidence* ou prevalence* ou new case?).tw,kw,kf.	3904190
6	4 et 5	32263
7	exp autism spectrum disorder/di, dg, ep ou Attention Deficit Disorder with Hyperactivity/di, dg, ep ou "Attention Deficit and Disruptive Behavior Disorders"/di, dg, ep ou Fetal Alcohol Spectrum Disorders/di, dg, ep	25017
8	6 ou 7 [Problèmes du développement]	44704
9	exp medical records/ ou international classification of diseases/ ou exp diagnostic techniques, neurological/ ou exp clinical laboratory techniques/ ou "diagnostic techniques and procedures"/	3171646
10	((patient* ou medical ou health) adj3 record*) ou ((diagnos* ou defin* ou classificat*) adj2 disease list) ou ((self report* ou standard*) adj2 measure*) ou (icd adj (cod* ou classification*)) ou (international adj3 classification)).tw,kw,kf.	287276
11	medicaid/ ou birth certificates/ ou death certificates/ ou hospital records/ ou insurance claim reporting/ ou exp insurance, health/ ou databases, factual/ ou information systems/ ou databases as topic/ ou database management systems/ ou software/ ou insurance claim review/ ou patient discharge/ ou exp registries/ ou utilization review/	519079
12	((administrat* ou physician* ou inpatient* ou emergency* ou hospital* ou clinic ou clinics ou pharmac* ou insurance) adj4 (admission* ou data ou dataset* ou database* ou data base? ou data bank? ou claim* ou billing* ou record* ou utilizat* ou utilisat*)) ou ((claim* ou discharg*) adj2 data*) ou ((database? ou databank? ou data base? ou data bank?) adj4 (factual ou administrat* ou claim? ou register* ou registr* ou topic? ou system?)) ou (claim? adj2 (analy* ou review? ou physician? ou pharmac* ou drug?)) ou (insurance adj2 (claim* ou audit*)) ou medicaid ou (health* adj2 plan) ou ((death* ou birth*) adj1 (certificate* ou record*)).tw,kw,kf. ou (database? ou databank? ou data base? ou data bank?).ti.	334466
13	ou/9-12 [Diagnostic]	4068131
14	8 et 13	6575
15	exp Diagnostic Errors/ ou Diagnosis, Differential/ ou "Predictive Value of Tests"/ ou "Sensitivity and Specificity"/ ou ROC Curve/ ou Area under Curve/ ou Bayes Theorem/ ou algorithms/ ou validation studies as topic/	1352991
16	((diagnos* adj3 (schedul* ou clinical* ou technique* ou procedur* ou assess* ou standard* ou error* ou false ou incorrect* ou wrong* ou correct*)) ou misdiagnos* ou (clinical* adj (standard* ou criter* ou measure* ou classif* ou technique* ou assess*))) ou ((positive ou negative) adj2 (predict* ou false)) ou sensitiv* ou specif* ou accura* ou valid* ou reliab* ou agree* ou concord* ou misclass* ou ((case ou cases) adj2 ascertain*) ou algorithm? ou (bayes* adj (theorem ou analysis ou approach ou forecast ou method ou prediction)) ou (roc adj (curve ou analysis)) ou receiver operating characteristic).tw,kw,kf.	6345372
17	15 ou 16	7034188
18	14 et 17	2895
19	limit 18 to (yr="1995-Current" et (english ou french))	2611
20	(202007* ou 202008* ou 202009* ou 20201* ou 202*).ez.	1831745
21	19 et 20	182

ANNEXE B
Évaluation de la qualité des rapports des études incluses à l'aide de la liste de vérification modifiée STARD^a aux fins de la validation des données administratives sur la santé

Section, sujet et point	Bickford, 2020 ²²	Brooks, 2021 ²³	Brooks, 2021 ²⁴	Burke, 2014 ²⁵	Coleman, 2015 ²⁶	Coo, 2017 ²⁷	Dodds, 2009 ²⁸	Hagberg, 2017 ²⁹	Lauritsen, 2010 ³⁰	Surén, 2019 ³¹	Daley, 2014 ³²	Gruschow, 2016 ³³	Mohr-Jensen, 2016 ³⁴	Morkem, 2020 ³⁵
TITRE, MOTS-CLÉS, RÉSUMÉ														
1. Identifie l'article comme une étude sur l'exactitude diagnostique?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
2. Identifie l'article comme une étude sur des données administratives?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non	Non	Oui	Oui	Non	Oui
INTRODUCTION														
3. Indique que l'identification et la validation de la maladie font partie des objectifs de l'étude?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
MÉTHODOLOGIE														
<i>Participants de la cohorte de validation</i>														
4. Décrit la cohorte de validation (cohorte de patients à laquelle le test de référence a été appliqué)?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
4a. Âge?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
4b. Maladie?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
4c. Gravité?	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non
4d. Lieu/région?	Oui	Oui	Oui	Non	Non	Oui	Oui	Non	Oui	Oui	Non	Oui	Oui	Non
5. Décrit la procédure de recrutement de la cohorte de validation?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
5a. Critères d'inclusion?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
5b. Critères d'exclusion?	Oui	Non	Non	Oui	Oui	Oui	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui
6. Décrit l'échantillonnage des patients (aléatoire, consécutif, tous, etc.)?	Oui	Oui	Oui	Oui	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
7. Décrit la collecte des données?	s.o.	Oui	Oui	Oui	Oui	Oui	s.o.	Oui	Oui	Oui	Oui	Oui	Oui	Oui
7a. Qui a identifié les patients et veillé à ce que leur sélection respecte les critères de recrutement des patients?	s.o.	Oui	Oui	Oui	Oui	Oui	s.o.	Oui	Oui	Oui	Oui	Oui	Oui	Oui
7b. Qui a recueilli les données?	s.o.	Oui	Oui	Oui	Oui	Oui	s.o.	Oui	Oui	Oui	Oui	Oui	Oui	Oui
7c. Formulaire de collecte de données a priori?	s.o.	Non	Oui	Oui	Oui	Oui	s.o.	Non	Incertain	Oui	Oui	Non	Oui	Non
7d. Comment la maladie a été classée?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non
8. Y avait-il un échantillon fractionné (c.-à-d. revalidation à l'aide d'une cohorte distincte)?	Non	Oui	Non ^b	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non

Suite à la page suivante

ANNEXE B (suite)
Évaluation de la qualité des rapports des études incluses à l'aide de la liste de vérification modifiée STARD^a aux fins de la validation des données administratives sur la santé

Section, sujet et point	Bick-ford, 2020 ²²	Brooks, 2021 ²³	Brooks, 2021 ²⁴	Burke, 2014 ²⁵	Coleman, 2015 ²⁶	Coo, 2017 ²⁷	Dodds, 2009 ²⁸	Hagberg, 2017 ²⁹	Lauritsen, 2010 ³⁰	Surén, 2019 ³¹	Daley, 2014 ³²	Gruschow, 2016 ³³	Mohr-Jensen, 2016 ³⁴	Morkem, 2020 ³⁵
Méthodes de tests														
9. Décrit le nombre de personnes ayant fait la lecture du test de référence, ainsi que la formation et l'expertise de ces personnes?	s.o.	Oui	Oui	Oui	Oui	Non	s.o.	Non	Oui	Oui	Oui	Oui	Oui	Non
10. Si plus d'une personne a fait la lecture du test de référence, indique la mesure de la cohérence (p. ex. kappa)?	s.o.	Non	Non	Oui	Oui	s.o.	s.o.	Non	Oui	s.o.	Non	Oui	Oui	s.o.
11. Les personnes ayant fait la lecture du test de référence (validation) ignoraient-elles les résultats de la classification des patients selon les données administratives? (p. ex. la personne qui a évalué les dossiers ignorait-elle les codes de facturation attribués dans les dossiers?)	s.o.	Oui	Oui	Incer-tain	Non	Oui	s.o.	Non	Non	Non	Non	Non	Non	Oui
Méthodes statistiques														
12. Décrit les méthodes de calcul ou de comparaison de l'exactitude diagnostique?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non	Oui
RÉSULTATS														
Participants														
13. Mentionne la période durant laquelle l'étude s'est déroulée et les dates de début et de fin du recrutement?	Oui	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non	Oui	Non
14. Décrit le nombre de personnes répondant aux critères d'inclusion ou d'exclusion?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
15. Diagramme de flux de l'étude?	Non	Oui	Oui	Oui	Non	Non	Non	Non	Non	Oui	Oui	Oui	Oui	Non
Résultats des tests														
16. Indique la répartition des différents degrés de gravité de la maladie?	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non
17. Contient un tableau croisé comparant les résultats des tests étudiés aux résultats du test de référence?	Oui	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non	Oui	Oui	Oui

Suite à la page suivante

ANNEXE B (suite)
Évaluation de la qualité des rapports des études incluses à l'aide de la liste de vérification modifiée STARD^a aux fins de la validation des données administratives sur la santé

Section, sujet et point	Bick-ford, 2020 ²²	Brooks, 2021 ²³	Brooks, 2021 ²⁴	Burke, 2014 ²⁵	Coleman, 2015 ²⁶	Coo, 2017 ²⁷	Dodds, 2009 ²⁸	Hagberg, 2017 ²⁹	Lauritsen, 2010 ³⁰	Surén, 2019 ³¹	Daley, 2014 ³²	Gruschow, 2016 ³³	Mohr-Jensen, 2016 ³⁴	Morkem, 2020 ³⁵
Estimations														
18. Présente au moins quatre estimations de l'exactitude diagnostique? (estimations présentées dans les études incluses)	Oui	Oui	Oui	Non	Non	Non	Non	Non	Non	Non	Non	Oui	Non	Non
18a. Sensibilité	Oui	Oui	Oui	Non	Non	Oui	Oui	Non	Non	Non	Non	Oui	Non	Non
18b. Spécificité	Oui	Oui	Oui	Non	Non	Non	Oui	Non	Non	Non	Non	Oui	Non	Non
18c. VPP	Oui	Oui	Oui	Oui	Oui	Oui	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui
18d. VPN	Oui	Oui	Oui	Oui	Non	Non	Non	Non	Non	Non	Non	Oui	Non	Oui
18e. Rapports de vraisemblance	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non
18f. Kappa	Oui	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Oui	Non	Non
18g. Aire sous la courbe ROC/statistique C	Oui	Non	Non	Non	Non	Non	Oui	Non	Non	Non	Non	Non	Non	Non
18h. Exactitude/concordance	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Oui	Non	Non
19. Les données sur l'exactitude sont-elles présentées pour des sous-groupes (p. ex. âge, situation géographique, sexe)?	Non	Non	Non	Oui	Oui	Oui	Oui	Non	Non	Oui	Non	Non	Non	Non
20. Si la VPP ou la VPN est calculée, est-ce que le ratio cas-témoins de la cohorte de validation se rapproche de la prévalence de la maladie dans la population?	Non	Oui	Oui	Non	Non	Non	s.o.	Non	Non	Non	Non	Non	Non	Non
21. Calcule l'IC à 95 % pour chaque mesure de l'exactitude diagnostique?	Oui	Oui	Oui	Non	Non	Oui	Non	Non	Oui	Oui	Oui	Oui	Non	Oui
ANALYSE														
22. Traite de l'applicabilité des résultats?	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui

Abréviations : IC, intervalle de confiance; ROC, Receiver Operating Characteristic; s.o., sans objet; VPN, valeur prédictive négative; VPP, valeur prédictive positive.

^a Standards for the Reporting of Diagnostic Accuracy Studies (STARD) : méthode d'évaluation de la qualité des rapports des études de validation faisant appel à des données administratives.

^b Cette étude revalidait un algorithme à l'aide d'une cohorte distincte; toutefois, comme l'algorithme n'utilisait pas de données administratives sur la santé, il ne faisait pas partie de la portée de la revue.

ANNEXE C
Évaluations du risque de biais et de l'applicabilité à l'aide de l'outil QUADAS-2^a

Domaines d'étude	Bickford, 2020 ²²	Brooks, 2021 ²³	Brooks, 2021 ²⁴	Burke, 2014 ²⁵	Coleman, 2015 ²⁶	Coo, 2017 ²⁷	Dodds, 2009 ²⁸	Hagberg, 2017 ²⁹	Lauritsen, 2010 ³⁰	Surén, 2019 ³¹	Daley, 2014 ³²	Gruschow, 2016 ³³	Mohr-Jensen, 2016 ³⁴	Morkem, 2020 ³⁵
1. SÉLECTION DES PATIENTS														
<i>A. Risque de biais</i>														
Q1	Oui	Oui	Oui	Non	Incertain	Non ^b Oui ^c	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non
Q2	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Q3	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non	Non	Non	Oui	Non
Risque	ÉLEVÉ	FAIBLE	FAIBLE	ÉLEVÉ	INCERTAIN	ÉLEVÉ ^b FAIBLE ^c	FAIBLE	FAIBLE	FAIBLE	ÉLEVÉ	ÉLEVÉ	FAIBLE	FAIBLE	ÉLEVÉ
<i>B. Préoccupations relatives à l'applicabilité</i>														
Préoccupation	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE
2. TEST ÉVALUÉ														
<i>A. Risque de biais</i>														
Q1	Incertain	Incertain	Incertain	Oui	Oui	Non	Incertain	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Risque	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE
<i>B. Préoccupations relatives à l'applicabilité</i>														
Préoccupation	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE
3. TEST DE RÉFÉRENCE														
<i>A. Risque de biais</i>														
Q1	Oui	Oui	Oui	Oui	Oui	Incertain ^b Non ^c	Oui	Incertain	Oui	Oui	Oui	Non	Oui	Incertain
Q2	Oui	Oui	Oui	Incertain	Non	Oui ^b Non ^c	Oui	Non	Non	Non	Non	Non	Non	Oui
Risque	FAIBLE	FAIBLE	FAIBLE	FAIBLE	INCERTAIN	FAIBLE	INCERTAIN ^b ÉLEVÉ ^c	FAIBLE	FAIBLE	FAIBLE	FAIBLE	ÉLEVÉ	FAIBLE	INCER-TAIN
<i>B. Préoccupations relatives à l'applicabilité</i>														
Préoccupation	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE
4. SUIVI ET ORGANISATION TEMPORELLE														
<i>A. Risque de biais</i>														
Q1	Oui	Incertain	Incertain	Oui	Incertain	Oui	Oui	Oui	Incertain	Oui	Oui	Oui	Oui	Incertain
Q2	Oui	Oui	Oui	Oui	Oui	Non	Oui	Oui	Oui	Oui	Oui	Non	Oui	Oui
Q3	Oui	Oui	Oui	Oui	Non	Oui	Non	Oui	Non	Non	Oui	Oui	Non	Oui
Risque	FAIBLE	FAIBLE	FAIBLE	FAIBLE	FAIBLE	ÉLEVÉ	ÉLEVÉ	FAIBLE	ÉLEVÉ	ÉLEVÉ	FAIBLE	ÉLEVÉ	ÉLEVÉ	FAIBLE

^a Outil révisé QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) : méthode d'évaluation du risque de biais et de l'applicabilité des études diagnostiques dans quatre domaines d'étude clés; pour chaque domaine, on attribue une cote au risque de biais (faible, élevé ou incertain) et à l'applicabilité de l'étude à la question de recherche (faible, élevé ou incertain).

^b Risque de biais pour la partie de l'étude qui évalue la sensibilité à l'aide de la « cohorte de sensibilité » de l'étude.

^c Risque de biais pour la partie de l'étude qui évalue la valeur prédictive positive en fonction des enfants ayant reçu un diagnostic administratif de trouble du spectre de l'autisme.

ANNEXE C (suite)
Évaluations du risque de biais et de l'applicabilité à l'aide de l'outil QUADAS-2^a

Risque de biais et préoccupations relatives à l'applicabilité : questions filtres et lignes directrices utilisées pour la cotation

Domaines d'étude

1. SÉLECTION DES PATIENTS

A. Risque de biais

Les patients ont-ils été recrutés de manière consécutive ou aléatoire?

Oui/Non/Incertain

Q1

- Sélectionner « Oui » si les patients de la cohorte de validation ont été sélectionnés de manière consécutive ou aléatoire.
- Sélectionner « Non » si la sélection des patients s'est faite de manière non consécutive ou par échantillonnage de commodité.
- Sélectionner « Incertain » si l'information fournie est insuffisante.

A-t-on évité le plan cas-témoin?

Oui/Non/Incertain

Q2

- Sélectionner « Oui » si l'on a évité le plan cas-témoin.
- Sélectionner « Non » si les patients ont été sélectionnés en fonction de leur statut malade (c.-à-d. cas confirmés plutôt que soupçonnés) ou non malade.
- Sélectionner « Incertain » si l'information fournie est insuffisante.

A-t-on évité les exclusions inappropriées?

Oui/Non/Incertain

Q3

- Sélectionner « Oui » si l'on a évité les exclusions inappropriées.
- Sélectionner « Non » si des patients ont fait l'objet d'exclusions inappropriées (p. ex. exclusion des patients dont le diagnostic est difficile à établir, ou dont le diagnostic est soupçonné, mais non confirmé).
- Sélectionner « Incertain » si l'information fournie est insuffisante.

Risque

La sélection des patients a-t-elle pu introduire un biais?

FAIBLE/ÉLEVÉ/INCERTAIN^a

B. Préoccupations relatives à l'applicabilité

Préoccupation

Craint-on que les patients sélectionnés ne correspondent pas à la question de recherche?

FAIBLE/ÉLEVÉE/INCERTAINE

2. TEST ÉTUDIÉ

A. Risque de biais

A-t-on interprété les résultats de l'algorithme de base de données administratives sans connaître les résultats du test de référence?

Oui/Non/Incertain

Q1

- Sélectionner « Oui » si l'on a interprété les résultats de l'algorithme sans connaître le diagnostic du test de référence.
- Sélectionner « Non » si l'on connaît le diagnostic du test de référence lors de l'interprétation des résultats de l'algorithme.
- Sélectionner « Incertain » si l'information fournie est insuffisante.

Risque

L'application ou l'interprétation des résultats de l'algorithme ont-elles pu introduire un biais?

FAIBLE/ÉLEVÉ/INCERTAIN^a

B. Préoccupations relatives à l'applicabilité

Préoccupation

Craint-on que l'algorithme, son application ou l'interprétation de ses résultats diffèrent de la question de recherche?

FAIBLE/ÉLEVÉE/INCERTAINE

3. TEST DE RÉFÉRENCE

A. Risque de biais

Le test de référence est-il susceptible de classer correctement le trouble visé?

Oui/Non/Incertain

Q1

- Sélectionner « Oui » si des critères de classification clinique établis, des définitions de cas clinique tirées des dossiers médicaux ou un diagnostic au dossier médical ont été utilisés, si du personnel chevronné ou dûment formé a procédé à l'abstraction ou à l'examen des données des dossiers (le cas échéant) et si la concordance calculée est élevée (dans les cas où plus d'une personne a procédé à l'abstraction ou à l'examen des données).
- Sélectionner « Non » si le test de référence est une mesure déclarée par le patient, si le personnel responsable de l'abstraction ou de l'examen des données tirées du test de référence n'était pas suffisamment chevronné ou formé (le cas échéant) ou si la concordance interévaluateurs est faible (dans les cas où plus d'une personne a procédé à l'abstraction ou à l'examen des données).
- Sélectionner « Incertain » si l'information fournie est insuffisante (p. ex. aucun renseignement fourni sur la concordance interévaluateurs dans les cas où l'examen des données a été effectué par plus d'une personne).

Suite à la page suivante

ANNEXE C (suite)
Évaluations du risque de biais et de l'applicabilité à l'aide de l'outil QUADAS-2^a

Domaines d'étude	
	A-t-on interprété les résultats du test de référence sans connaître les résultats de l'algorithme? Oui/Non/Incertain
Q2	<ul style="list-style-type: none"> • Sélectionner « Oui » si l'on a interprété les résultats du test de référence sans connaître les résultats de l'algorithme. • Sélectionner « Non » si l'on connaissait les résultats de l'algorithme lors de l'interprétation des résultats du test de référence, y compris dans les cas où seuls les patients repérés par l'algorithme étaient soumis au test de référence. • Sélectionner « Incertain » si l'information fournie est insuffisante.
Risque	
	Le test de référence, sa réalisation ou l'interprétation de ses résultats pourraient-ils avoir introduit un biais? FAIBLE/ÉLEVÉ/INCERTAIN ^a
B. Préoccupations relatives à l'applicabilité	
Préoccupation	Craint-on que le trouble visé par le test de référence diffère de celui de la question de recherche? FAIBLE/ÉLEVÉE/INCERTAINE
4. SUIVI ET ORGANISATION TEMPORELLE	
	Risque de biais
	Le délai entre la détermination des cas à l'aide de l'algorithme et le test de référence est-il approprié? Oui/Non/Incertain
Q1	<ul style="list-style-type: none"> • Sélectionner « Oui » si le délai entre l'application de l'algorithme et le test de référence est approprié. • Sélectionner « Non » si le délai entre le diagnostic du test de référence et le diagnostic de l'algorithme n'est pas approprié. • Sélectionner « Incertain » si l'information fournie est insuffisante.
	Les patients ont-ils tous passé le même test de référence? Oui/Non/Incertain
Q2	<ul style="list-style-type: none"> • Sélectionner « Oui » si les patients ont tous passé le même test de référence. • Sélectionner « Non » si des tests de référence différents ont été utilisés. • Sélectionner « Incertain » si l'information fournie est insuffisante.
	Les patients ont-ils tous été inclus dans l'analyse? Oui/Non/Incertain
Q3	<ul style="list-style-type: none"> • Sélectionner « Oui » si le nombre de patients inscrits (c.-à-d. le nombre de patients restants après les exclusions) est le même que le nombre de patients inclus dans le tableau à double entrée des résultats. • Sélectionner « Non » si le nombre de patients inscrits diffère du nombre de patients inclus dans le tableau à double entrée des résultats. • Sélectionner « Incertain » si l'information fournie est insuffisante (p. ex. aucun renseignement fourni sur la méthode utilisée pour constituer la population finale de l'étude de validation).
	Risque
	Le suivi des patients a-t-il pu introduire un biais? FAIBLE/ÉLEVÉ/INCERTAIN ^a

^a Lignes directrices utilisées pour la cotation :

- Lorsque la réponse à toutes les questions filtres d'un domaine était « Oui », le risque de biais était jugé « FAIBLE ».
- Lorsque la réponse à toutes les questions filtres d'un domaine était « Non », le risque de biais était jugé « ÉLEVÉ ».
- Lorsque la réponse à toutes les questions filtres d'un domaine était « Incertain », le risque de biais était jugé « INCERTAIN ».
- Lorsque la réponse à l'une des questions filtres était « Non », cela évoquait un potentiel de biais; les auteures de la revue ont décidé des critères en fonction desquels le risque de biais serait jugé élevé dans de telles circonstances.
 - ▶ La question filtre (Q1) du domaine « Test évalué » a été considérée comme une source moins importante de biais aux fins de la présente revue. La deuxième question filtre (Q2) du domaine « Test de référence » a elle aussi été considérée comme une source moins importante de biais, mais les auteures ont porté un jugement en fonction de chaque étude. Pour toutes les autres questions filtres, le fait de répondre « Non » à la question était suffisant pour que le risque de biais soit jugé élevé.
- Lorsque la réponse à l'une des questions filtres était « Incertain », cela évoquait un *potentiel* de biais; les auteures de la revue ont décidé des critères en fonction desquels le risque de biais serait jugé incertain dans de telles circonstances.

Références

1. American Psychiatric Association. DSM-5 : manuel diagnostique et statistique des troubles mentaux. 5^e édition. Issy-les-Moulineaux (France) : Elsevier Masson; 2015.
2. Langlois K, Samokhvalov A, Rehm J, Spence S, Conner Gorber S. Descriptions des états de santé au Canada : maladies mentales. Ottawa (Ontario) : Statistique Canada; 2011 [Statistique Canada, n° 82-619-MIE2005002 au catalogue].
3. Chudley AE, Conry J, Cook JL, Loock C, Rosales T, LeBlanc N. Ensemble des troubles causés par l'alcoolisation fœtale : lignes directrices canadiennes concernant le diagnostic. CMAJ 2005; 172(5 suppl.):SF1-SF21. <https://doi.org/10.1503/cmaj.050098>
4. Shaw M, Hodgkins P, Caci H, et al. A systematic review and analysis of long-term outcomes in attention deficit hyperactivity disorder: effects of treatment and non-treatment. BMC Med. 2012;10:99. <https://doi.org/10.1186/1741-7015-10-99>
5. Elder JH, Kreider CM, Brasher SN, Ansell M. Clinical impact of early diagnosis of autism on the prognosis and parent-child relationships. Psychol Res Behav Manag. 2017;10:283-292. <https://doi.org/10.2147/PRBM.S117499>
6. Blais C, Dai S, Waters C, et al. Assessing the burden of hospitalized and community-care heart failure in Canada. Can J Cardiol. 2014;30(3): 352-358. <https://doi.org/10.1016/j.cjca.2013.12.013>
7. O'Donnell S, Canadian Chronic Disease Surveillance System (CCDSS) Osteoporosis Working Group. Use of administrative data for national surveillance of osteoporosis and related fractures in Canada: results from a feasibility study. Arch Osteoporos. 2013;8:143 <https://doi.org/10.1007/s11657-013-0143-2>
8. Robitaille C, Dai S, Waters C, et al. Diagnosed hypertension in Canada: incidence, prevalence and associated mortality. CMAJ. 2012;184(1):E49-E56. <https://doi.org/10.1503/cmaj.101863>
9. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005; 58(4):323-337. <https://doi.org/10.1016/j.jclinepi.2004.10.012>
10. Goldfield N, Villani J. The use of administrative data as the first step in the continuous quality improvement process. Am J Med Qual. 1996;11(1): S35-S38.
11. Schwartz RM, Gagnon DE, Muri JH, Zhao QR, Kellogg R. Administrative data for quality improvement. Pediatrics. 1999;103(1, Suppl E):291-301.
12. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. J Clin Epidemiol. 2011; 64(8):821-829. <https://doi.org/10.1016/j.jclinepi.2010.10.006>
13. Lix L, Yogendran M, Burchill C, et al. Defining and validating chronic diseases: an administrative data approach. Winnipeg (MB): Manitoba Centre for Health Policy; 2006.
14. Sørensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. Int J Epidemiol. 1996;25(2): 435-442. <https://doi.org/10.1093/ije/25.2.435>
15. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Clin Chem. 2015; 61(12):1446-1452. <https://doi.org/10.1373/clinchem.2015.246280>
16. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6(11):e012799. <https://doi.org/10.1136/bmjopen-2016-012799>
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. <https://doi.org/10.1136/bmj.n71>
18. Spasoff R. Epidemiologic methods for health policy. 1st ed. New York (NY): Oxford University Press; 1999.
19. Shortliffe E, Cimino J, editors. Biomedical informatics: computer applications in health care and biomedicine. 3^e éd. New York (NY): Springer; 2006.
20. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-536. <https://doi.org/10.7326/0003-4819-155-8-20110180-00009>
21. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York (UK): University of York; 2009.
22. Bickford CD, Oberlander TF, Lanphear NE, et al. Identification of pediatric autism spectrum disorder cases using health administrative data. Autism Res. 2020;13(3):456-463. <https://doi.org/10.1002/aur.2252>
23. Brooks JD, Arneja J, Fu L, et al. Assessing the validity of administrative health data for the identification of children and youth with autism spectrum disorder in Ontario. Autism Res. 2021;14(5):1037-1045. <https://doi.org/10.1002/aur.2491>
24. Brooks JD, Bronskill SE, Fu L, et al. Identifying children and youth with autism spectrum disorder in electronic medical records: examining health system utilization and comorbidities. Autism Res. 2021;14(2):400-410. <https://doi.org/10.1002/aur.2419>
25. Burke JP, Jain A, Yang W, et al. Does a claims diagnosis of autism mean a true case? Autism. 2014;18(3):321-330. <https://doi.org/10.1177/1362361312467709>
26. Coleman KJ, Lutsky MA, Yau V, et al. Validation of autism spectrum disorder diagnoses in large healthcare systems with electronic medical records. J Autism Dev Disord. 2015;45(7):1989-1996. <https://doi.org/10.1007/s10803-015-2358-0>

27. Coo H, Ouellette-Kuntz H, Brownell M, Shooshtari S, Hanlon-Dearman A. Validating an administrative data-based case definition for identifying children and youth with autism spectrum disorder for surveillance purposes. *Can J Public Health*. 2017; 108(5-6):e530-e538. <https://doi.org/10.17269/cjph.108.5963>
28. Dodds L, Spencer A, Shea S, et al. Validité des diagnostics d'autisme recensés à l'aide de données administratives sur la santé. *Maladies chroniques au Canada*. 2009;29(3):114-120.
29. Hagberg KW, Jick SS. Validation of autism spectrum disorder diagnoses recorded in the Clinical Practice Research Datalink, 1990-2014. *Clin Epidemiol*. 2017;9:475-482. <https://doi.org/10.2147/CLEP.S139107>
30. Lauritsen MB, Jørgensen M, Madsen KM, et al. Validity of childhood autism in the Danish psychiatric central register: findings from a cohort sample born 1990-1999. *J Autism Dev Disord*. 2010;40(2):139-148. <https://doi.org/10.1007/s10803-009-0818-0>
31. Surén P, Havdahl A, Øyen AS, et al. Diagnosing autism spectrum disorder among children in Norway. *Tidsskr Nor Laegeforen*. 2019;139(14):10.4045/tidsskr.18.0960. <https://doi.org/10.4045/tidsskr.18.0960>
32. Daley MF, Newton DA, Debar L, et al. Accuracy of electronic health record-derived data for the identification of incident ADHD. *J Atten Disord*. 2014; 21(5):416-425. <https://doi.org/10.1177/1087054713520616>
33. Gruschow SM, Yerys BE, Power TJ, Durbin DR, Curry AE. Validation of the use of electronic health records for classification of ADHD status. *J Atten Disord*. 2016;23(13):1647-1655. <https://doi.org/10.1177/1087054716672337>
34. Mohr-Jensen C, Vinkel Koch S, Briciet Lauritsen M, Steinhausen H-C. The validity and reliability of the diagnosis of hyperkinetic disorders in the Danish Psychiatric Central Research Registry. *Eur Psychiatry*. 2016;35:16-24. <https://doi.org/10.1016/j.eurpsy.2016.01.2427>
35. Morkem R, Handelman K, Queenan JA, Birtwhistle R, Barber D. Validation of an EMR algorithm to measure the prevalence of ADHD in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN). *BMC Med Inform Decis Mak*. 2020;20:166. <https://doi.org/10.1186/s12911-020-01182-2>
36. Thanh NX, Jonsson E, Salmon A, Sebastianski M. Incidence and prevalence of fetal alcohol spectrum disorder by sex and age group in Alberta, Canada. *J Popul Ther Clin Pharmacol*. 2014;21(3):e395-e404.
37. Clarren SK, Lutke J, Sherbuck M. The Canadian guidelines and the interdisciplinary clinical capacity of Canada to diagnose fetal alcohol spectrum disorder. *J Popul Ther Clin Pharmacol*. 2011;18(3):e494-e499.
38. Cook JL, Green CR, Lilley CM, et al. Fetal alcohol spectrum disorder: a guideline for diagnosis across the lifespan. *CMAJ*. 2016;188(3):191-197. <https://doi.org/10.1503/cmaj.141593>
39. Brown JM, Bland R, Jonsson E, Greenshaw AJ. The standardization of diagnostic criteria for fetal alcohol spectrum disorder (FASD): implications for research, clinical practice and population health. *Can J Psychiatry*. 2019;64(3):169-176. <https://doi.org/10.1177/0706743718777398>
40. Widdifield J, Labrecque J, Lix L, et al. Systematic review and critical appraisal of validation studies to identify rheumatic diseases in health administrative databases. *Arthritis Care Res (Hoboken)*. 2013;65(9):1490-1503. <https://doi.org/10.1002/acr.21993>
41. Morris-Rosendahl D, Crocq M-A. Neurodevelopmental disorders—the history and future of a diagnostic concept. *Dialogues Clin Neurosci*. 2020;22(1):65-72. <https://doi.org/10.31887/DCNS.2020.22.1/macrocq>
42. Stewart L, Moher D, Shekelle P. Why prospective registration of systematic reviews makes sense. *Syst Rev*. 2012; 1:7. <https://doi.org/10.1186/2046-4053-1-7>