Evidence synthesis

Accuracy of administrative database algorithms for autism spectrum disorder, attention-deficit/hyperactivity disorder and fetal alcohol spectrum disorder case ascertainment: a systematic review

Siobhan O'Donnell, MSc; Sarah Palmeter, MPH; Meghan Laverty, MSc; Claudia Lagacé, MSc

This article has been peer reviewed.

🐦 Tweet this article

Abstract

Introduction: The purpose of this study was to perform a systematic review to assess the validity of administrative database algorithms used to identify cases of autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD) and fetal alcohol spectrum disorder (FASD).

Methods: MEDLINE, Embase, Global Health and PsycInfo were searched for studies that validated algorithms for the identification of ASD, ADHD and FASD in administrative databases published between 1995 and 2021 in English or French. The grey literature and reference lists of included studies were also searched. Two reviewers independently screened the literature, extracted relevant information, conducted reporting quality, risk of bias and applicability assessments, and synthesized the evidence qualitatively. PROSPERO CRD42019146941.

Results: Out of 48 articles assessed at full-text level, 14 were included in the review. No studies were found for FASD. Despite potential sources of bias and significant between-study heterogeneity, results suggested that increasing the number of ASD diagnostic codes required from a single data source increased specificity and positive predictive value at the expense of sensitivity. The best-performing algorithms for the identification of ASD were based on a combination of data sources, with physician claims database being the single best source. One study found that education data might improve the identification of ASD (i.e. higher sensitivity) in school-aged children when combined with physician claims data; however, additional studies including cases without ASD are required to fully evaluate the diagnostic accuracy of such algorithms. For ADHD, there was not enough information to assess the impact of number of diagnostic codes or additional data sources on algorithm accuracy.

Conclusion: There is some evidence to suggest that cases of ASD and ADHD can be identified using administrative data; however, studies that assessed the ability of algorithms to discriminate reliably between cases with and without the condition of interest were lacking. No evidence exists for FASD. Methodologically higher-quality studies are needed to understand the full potential of using administrative data for the identification of these conditions.

Keywords: autism spectrum disorder, attention deficit disorder with hyperactivity, fetal alcohol spectrum disorders, algorithms, validation study, administrative data, public health surveillance

Highlights

- Few studies have validated administrative database algorithms for the identification of ASD and ADHD. No validation studies were found for FASD.
- Extensive heterogeneity in study design and conduct across the included studies precluded a quantitative synthesis of the results.
- There is evidence to suggest that ASD and ADHD can be identified using administrative data; however, studies that assessed the ability of algorithms to discriminate reliably between cases with and without the condition of interest were lacking.
- The best-performing algorithms used to identify ASD are based on a combination of administrative data sources, with physician claims data being the single best source.
- Higher-quality studies are essential to fully leverage administrative data for surveillance and research on these conditions.

Introduction

Neurodevelopmental disorders, a group of conditions with onset early in life, are characterized by impairments in physical development, learning, language and/or behaviour.[1] Despite the wide-ranging personal and societal impacts that these

Author references:

Public Health Agency of Canada, Ottawa, Ontario, Canada

Correspondence: Siobhan O'Donnell, Centre for Surveillance and Applied Research, Health Promotion and Chronic Disease Prevention Branch, Public Health Agency of Canada, 785 Carling Avenue, AL 6806A, Ottawa, ON K1A 0K9; Tel: 613-301-7325; Email: siobhan.odonnell@phac-aspc.gc.ca

disorders have, early detection and interventions have been shown to improve outcomes in those with certain types of neurodevelopmental disorders, including autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD) and fetal alcohol spectrum disorder (FASD).[2-5] In light of this, a better understanding of the epidemiological burden of these disorders in Canada is essential in the implementation of public policy, including the establishment of programs and services.

Population-based administrative databases, designed for health system management and physician remuneration, offer an efficient and inexpensive way of providing longitudinal epidemiological data. As a result, these data are being increasingly used as a way to conduct chronic disease surveillance,[6-8] disease and treatment outcome research[9] and quality of care studies.[10,11] However, along with these advantages, health administrative databases have limitations, including the potential for misclassification.[12]

The accuracy of the diagnostic codes or their combination (algorithm) for surveillance or research purposes[13] depends on multiple factors, including database quality, the specific condition being identified and the validity of the diagnostic codes within the patient group.[12] Therefore, validation studies are necessary to evaluate the accuracy of algorithms used for case ascertainment.[14] Validation involves quantifying the number of instances in which the algorithm matches a reference standard, such as a medical record diagnosis.[12] In this way, the algorithm can be treated like a diagnostic test, and measures of diagnostic accuracy can be calculated. The results of these validation studies are typically reported as estimates of the sensitivity and specificity of the algorithm, which express how good the algorithm is at correctly identifying individuals with and without the target condition, respectively.[15,16] Other diagnostic accuracy statistics can be used, including positive predictive value (PPV) and negative predictive value (NPV).

To our knowledge, there are no published reviews that have evaluated the validity of health administrative database algorithms for the surveillance or research of neurodevelopmental disorders, specifically ASD, ADHD and FASD. Thus, the primary objective of this systematic review was to address this shortcoming. The secondary objective was to examine the impact of linking health to non-health (i.e. education or social services) administrative data on the accuracy of these algorithms.

Methods

This systematic review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[17] Ethics approval was not required, as primary data were not collected.

Protocol and registration

The protocol for this systematic review has been registered in the International Prospective Register of Systematic Reviews (PROSPERO 2019 CRD42019146941), was published on 16 December 2019 and is available from https://www.crd.york.ac .uk/prospero/display_record.php?ID =CRD42019146941.

Search strategy

A systematic search of MEDLINE, Embase, Global Health and PsycInfo was conducted to identify all validation studies using administrative data to ascertain cases of ASD, ADHD or FASD published in English or French from January 1995 to March 2021. A start year of 1995 was chosen for the database searches in order to align with the start year for data collection in the Canadian Chronic Disease Surveillance System, a collaborative network of provincial and territorial health administrative surveillance systems supported by the Public Health Agency of Canada. A reference librarian developed the search strategy using medical subject headings and keywords related to the target conditions (e.g. "exp autism spectrum disorder/"), administrative data (e.g. "exp insurance, health/"), reference standard (e.g. "exp medical records/") and validation testing (e.g. "sensitivity and specificity/"). The initial search strategy was developed in MEDLINE and was peer reviewed before being adopted for the other databases (Appendix A). Additionally, the grey literature was searched via two mechanisms: an advanced Google search and searching websites of relevant agencies and organizations. Furthermore, reference lists of relevant surveillance reports found in the grey literature as well as articles that met the eligibility criteria of the review were manually searched for additional studies.

Eligibility criteria

To be included, studies of any design type had to report

- the assessment or validation of one or more health administrative database algorithms against a reference standard (i.e. established clinical criteria, medical record diagnosis, electronic medical record or patient self-report measure) for identifying a case with ASD, ADHD or FASD; and

- at least one measure of diagnostic accuracy (i.e. sensitivity, specificity, PPV, NPV, area under the receiver operating characteristic curve [or C-statistic], Youden's index, kappa statistic or likelihood ratio).

An administrative database algorithm was defined as a set of rules for identifying disease cases from administrative data, with elements including type of data source, number of years of administrative data, diagnostic or medication code(s) and number of administrative data records (i.e. contacts) with diagnostic or medication code(s).[13] While the administrative database algorithm had to include health administrative data, it could also include other types of administrative data, such as education or social services data.

These algorithms could be based on administrative data from either a health administrative database or a clinical or health information system. A health administrative database was defined as information that is routinely or passively collected solely for administrative purposes in managing the health care of patients,[18] and a clinical/health information system was defined as administrative data supplemented with detailed clinical information by way of electronic health records.[19]

Abstracts, editorials and commentaries were excluded from the review, as well as studies published before 1995 or in a language other than English or French.

Study selection and data extraction

Two reviewers (CL and SO) independently screened the titles and abstracts of all bibliographic records and articles identified through electronic database searches, grey literature and reference lists of surveillance

reports for eligibility. When consensus could not be reached on a given study, it was retained for the next stage of screening. For every study that passed the title and abstract level of screening, full-text articles were assessed for eligibility by two reviewers (ML and SO) independently and the reason for exclusion was recorded. When reviewers did not agree on the inclusion or exclusion of an article, a third reviewer (CL) was consulted. The reference lists of all articles that passed full-text review were manually searched using the same two level screening process conducted by two reviewers (SO and SP).

Relevant information was extracted from included articles using a template developed for this systematic review and piloted before use that included author, year, geographic location, study cohort, type of administrative data source(s), administrative database algorithm(s) and related elements, reference standard, reference diagnostic criteria and measures of diagnostic accuracy. One reviewer (ML) completed the extraction, which was verified by a second (SO). Any disagreement was resolved by consensus, or when required, by a third party (CL).

Reporting quality, risk of bias and applicability assessments

Included studies underwent a reporting quality assessment using the 40-point, modified Standards for the Reporting of Diagnostic Accuracy Studies (STARD) checklist[12] (Appendix B) and risk of bias and applicability assessments using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool[20] (Appendix C). These assessments were completed by one reviewer (ML) and verified by a second (SO). Any disagreements were resolved by consensus, or if necessary, by a third reviewer (CL).

Data synthesis and analysis

Extensive heterogeneity in study design and conduct across the included studies precluded a quantitative synthesis; therefore, results were synthesized narratively using text and tables for the conditions of interest. Findings both within and between studies were explored as per guidance from the Centre for Reviews and Dissemination.[21] While the diagnostic accuracy of the administrative database algorithms in all included studies were considered, our final recommendations

also took the reporting quality, risk of bias and applicability concerns of each study into account.

Results

Search results

The PRISMA flow diagram in Figure 1 documents the study screening process.[17] A total of 5918 records were identified through database searching and 11 additional records through other sources (grey literature and surveillance report reference lists). After duplicates were removed, 4133 records identified from database searching were screened (title and abstract), of which 4085 were deemed ineligible and excluded. Of the remaining 48 records that underwent a full-text review for eligibility, 34 were excluded (17 did not use a health administrative database, 6 used a health administrative database, but were not validation studies, 8 did not validate a health administrative database algorithm, and 3 were excluded for other reasons) and 14 records (studies) were included in the review. None of the 11 records identified through grey literature and surveillance report reference lists were included in the review. No additional studies were found from manually searching reference lists of included articles.

Characteristics of included studies

The characteristics of the 14 included studies[22-35] are provided in Table 1. Ten studies focussed on ASD[22-31] and the remaining four on ADHD.[32-35] There were no studies identified for FASD.

ASD studies

Studies that validated algorithms to identify ASD were published between 2009[28] and 2021.[23,24] Five studies were performed in Canada,[22-24,27,28] two in the United States,[25,26] one in the United Kingdom,[29] one in Denmark[30] and one in Norway.[31] All 10 studies included children and youth as their study population,[22-31] although only seven reported the age range.[22-27,31]

Validation cohort sample sizes ranged from 37[29] to 10 000.[23,24] Patients were initially selected from diagnostic codes in the administrative database for five studies[25,26,29-31] and for one of the two samples used in one study.[27] Only five studies included a comparator group without ASD.[22-25,28] The prevalence of ASD in the

validation cohort ranged from 1.1%[23,24] to 67.9%.[22]

Six studies used health administrative databases,[22,23,25,28,30,31] two used a clinical/health information system,[24,29] one used both health administrative databases and a clinical/health information system[26] and one used health administrative databases combined with an education data source.[27] The most common data source included a combination of outpatient and inpatient data.[22,23,28,30,31]

A variety of diagnostic codes were used: International Classification of Diseases, Eighth Revision (ICD-8),[30] International Classification of Diseases, Ninth Revision (ICD-9),[22,23,25-28] International Classification of Diseases, Tenth Revision (ICD-10),[22,23,27,28,30,31] Ontario Health Insurance Plan physician billing codes,[23,24] Read codes[29] and unique codes for education and mental health services.[27] The number of algorithms validated within each study ranged from 1[29-31] to 153.[23]

Several reference standards were used, with the most common being a medical chart diagnosis.[23,24,27,29]

The PPV was the most commonly reported measure and was reported in 9 of the 10 studies.[22-27,29-31] Only three studies reported at least four measures of diagnostic accuracy.[22-24]

ADHD studies

Studies that validated algorithms to identify ADHD were published between 2014[32] and 2020.[35] Two studies were performed in the USA,[32,33] one in Canada[35] and one in Denmark.[34] Three studies included children and youth as their study population,[32-34] two of which reported the age range.[32,34] One study included adults and children aged four years and older.[35]

Validation cohort sample sizes ranged from 372[34] to 2837.[33] Patients were initially selected from diagnostic codes in the administrative data source for three studies[32-34] and from diagnostic codes and medication prescriptions in the administrative data source for one study.[35] Only two studies included a comparator group without ADHD.[33,35] The prevalence of ADHD in the validation cohort ranged from 50.0%[35] to 56.7%.[33]

FIGURE 1
PRISMA flow diagram
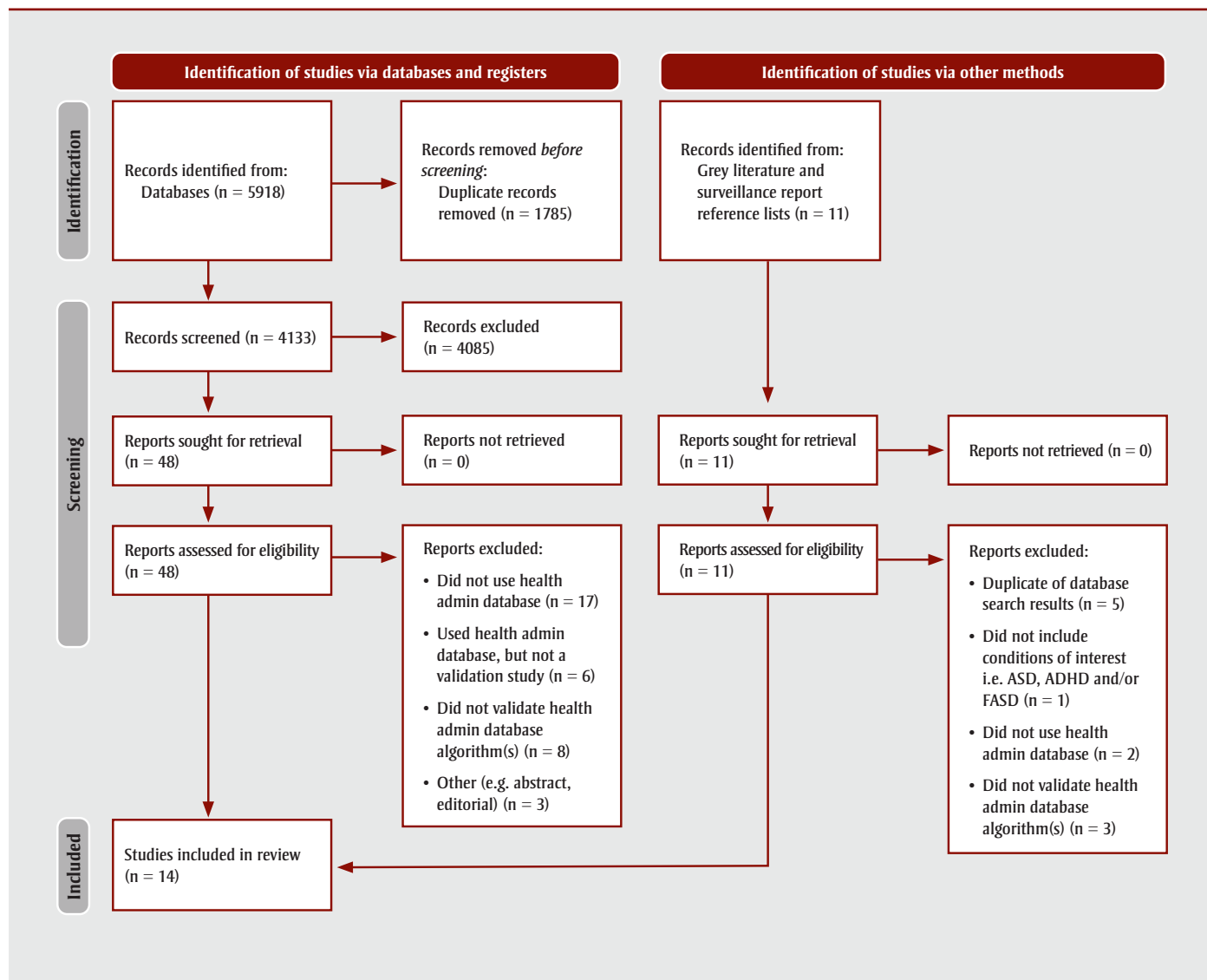
Note: PRISMA template from Page MJ et al.[17]

One study used a health administrative database, including inpatient and outpatient psychiatric hospital data[34] and three studies used a clinical/health information system, specifically, electronic health records.[32,33,35]

Two studies used ICD-9 codes,[32,33] one used ICD-10 codes,[34] and one used ICD-9 codes and medication prescriptions.[35] Each study validated one algorithm only.[32-35] One study captured incident, rather than prevalent, cases of ADHD.[32]

Various reference standards were used. One study used clinical classification criteria documented in the medical chart.[34] One used a medical chart ADHD diagnosis.[35] One used a clinical case definition

that required a combination of evidence from the electronic health record, and in the absence of this evidence, a manual review of the electronic health record.[33] Lastly, one used a combination of clinical classification criteria, medical record diagnosis and standardized screening checklist documented in the medical chart.[32]

The PPV was reported in all four studies,[32-35] while only one reported at least four measures of diagnostic accuracy.[33]

Reporting quality of included studies

The number and percentage of included studies meeting reporting criteria using the modified STARD checklist for validating health administrative data are summarized

in Table 2.[12] The quality of reporting was variable. Highlighted below are areas where the reporting quality was especially suboptimal, that is, where less than half of the studies met the criterion. For full details of the reporting quality results for each included study, see Appendix B.

ASD studies

Concerning the methods used, none of the ASD studies described the severity of the patients, only one re-validated the algorithms using a separate cohort[23] and just three of eight that included reviewers of the reference standard reported that the reviewers were blinded to the patient classification by administrative data.[23,24,27] In terms of the results, only four included a study flow diagram,[23-25,31] none reported

TABLE 1
Characteristics of included studies

First author, year Country	Validation population Age	Sample size	Administrative data source(s)	Years of administrative data	Diagnostic codes included in algorithm(s)
ASD					
Bickford, 2020[22] Canada	Children aged 1 to 14 years, born in British Columbia between 1 April 2000 and 31 December 2009, assessed in one of the British Columbia Autism Assessment Network centres between 1 April 2004 and 31 December 2014 or with a Ministry of Education designation of ASD between 1 September 2004 and 30 June 2015. 1–14 years.	8670 (cases and non-cases)	Health administrative databases[a]: hospital discharge abstracts and physician claims	2000–2014	Hospital discharge abstracts: ICD-9 299.x; ICD-10-CA F84.x Physician claims: ICD-9 299.x
Brooks, 2021[23] Canada	Children and youth aged 1 to 24 as of 31 December 2011, within the Electronic Medical Record Primary Care database (from over 350 Ontario family physicians) with a valid date of birth, registered with an active/practising physician who has used EMR for more than 2 years, alive as of load date, and present in EMR for at least 1 year. 1–24 years.	10 000 (cases and non-cases)	Health administrative databases[a]: hospital discharge abstracts, emergency department visits, outpatient surgery, physician claims	NR	Hospital discharge abstracts/ emergency department visits/ outpatient surgery: ICD-9 299.x; ICD-10-CA F84.x Physician claims: OHIP physician billing code 299
Brooks, 2021[24,b] Canada	Children and youth aged 1 to 24 as of 31 December 2011, within the Electronic Medical Record Primary Care database (from over 350 Ontario family physicians) with a valid date of birth, registered with an active/practising physician who has used EMR for more than 2 years, alive as of load date, and present in EMR for at least 1 year. 1–24 years.	10 000 (cases and non-cases)	Clinical/health information system[c]: electronic health records	NR	OHIP physician billing codes 299, 315
Burke, 2014[25] USA	Children and youth aged 2 to 20 years at time of first ASD or ASD-associated claim, insured through a large national private health plan. Eligible cases had to have at least 6 months of continuous enrolment pre- and post-first ASD or ASD-associated claim and not have any claims with a diagnosis of childhood disintegrative disorder or Rett's syndrome. 2–20 years.	432 (cases and non-cases)	Health administrative database[a]: private medical, pharmacy, and behavioural insurance claims	2001–2009	ASD: ICD-9 299.00–299.01, 299.80–299.81, 299.9 (in any position) ASD-associated conditions: ICD-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] USA	Children and youth aged < 18 years with current membership in one of the participating health care plans as of December 2010, with at least one ASD diagnostic code, that were not diagnosed in a specialty ASD centre. < 18 years.	1272 (cases only)	Health administrative database[a] and clinical/health information system[c]: insurance claims and electronic health records	1995–2010	ICD-9 299.0, 299.9, 299.8

Continued on the following page

TABLE 1 (continued)
Characteristics of included studies

First author, year Country	Validation population Age	Sample size	Administrative data source(s)	Years of administrative data	Diagnostic codes included in algorithm(s)
Coo, 2017[27] Canada	Children aged 2–14 years, born between 1997 and 2009 with an administrative diagnosis of ASD and/or who were confirmed as a case by a Manitoba child/youth behavioural or disability service provider on or before 31 December 2011. 2–14 years.	2610 (cases only)	Health administrative databases[a] and education database: hospital discharge abstracts, physician claims, mental health services and education data	1997–2011	Hospital discharge abstracts: ICD-9-CM 299.0, 299.8, 299.9; ICD-10-CA F84.0, F84.1, F84.5, F84.8, F84.9 (in any diagnostic field) Physician claims: ICD-9-CM 299.x ("most responsible" diagnosis) Education data: CATEGORYN=ASD (child received funding under special needs category for ASD) Mental health services: NDC-A 312 0.92 (enrolment in autism treatment program)
Dodds, 2009[28] Canada	Children born between 1989 and 2002, and assessed for ASD by a team of specialists between 2001 and 2005. Age not provided.	264 (cases and non-cases)	Health administrative databases[a]: hospital discharge abstracts, physician claims and mental health outpatient data	1989–2005	ICD-9 299.x or ICD-10 F84.x (primary or secondary diagnostic field)
Hagberg, 2017[29] United Kingdom	Singleton children born between 1990 and 2011, with at least three years of follow-up from birth. Age not provided.	37 (cases only)	Clinical/health information system[c]: electronic health records	1990–2014	Read codes: E140.00, E140000, E140100, E140.12, E140.13, E140z00, Eu84000, Eu84011, Eu84012, Eu84100, Eu84z11, Eu84500, Eu84.00, Eu84y00, Eu84z00
Lauritsen, 2010[30] Denmark	Children born between 1990 and 1999, whose parent(s) or legal guardian(s) resided in Denmark, with a reported diagnosis of childhood autism. Age not provided.	499 (cases only)	Health administrative database[a]: psychiatric inpatient and outpatient data	1990–2001	ICD-8 299.00 or ICD-10 F84.0 (main or subsidiary diagnosis)
Surén, 2019[31] Norway	Children born 1999–2009, enrolled in the Norwegian Mother, Father and Child Cohort Study, with a reported autism diagnosis in Norwegian Patient Registry between 2008 and 2014, aged 5–15 years at end of follow-up, with patient records available and who did not undergo a clinical assessment as part of the Autism Study. 5–15 years.	553 (cases only)	Health administrative database[a]: mental health care provider, somatic hospital, and specialist private consultant data	2008–2014	ICD-10 F84.x
ADHD					
Daley, 2014[32] USA	Children aged 3–9 years at time of first diagnosis, insured at one of eight managed care organizations or who sought care at one of two community health sites between 2004 and 2010, who met the case definition for incident ADHD and were without a diagnosis of mental retardation or pervasive developmental disorder. 3–9 years.	500 (cases only)	Clinical/health information system[c]: electronic health records	2004–2010	ICD-9-CM 314.0x

TABLE 1 (continued)

Characteristics of included studies

First author, year Country	Validation population Age	Sample size	Administrative data source(s)	Years of administrative data	Diagnostic codes included in algorithm(s)
Gruschow, 2016[33] USA	Patients of the Children's Hospital of Philadelphia health care network, born between 1987 and 1995 (median age 17.9 years) with ≥ 2 visits and who were New Jersey residents at the time of their last visit, that were not identified as having an intellectual disability and had their last visit at ≥ 12 years of age. Children with a recorded ADHD diagnosis in their electronic health record vs. children without were identified. Median age (IQR): 17.9 (15.9–19.1) years	2030 (cases) 807 (non-cases)	Clinical/health information system[c]: electronic health records	2001+	ICD-9-CM 314.x
Mohr-Jensen, 2016[34] Denmark	Children and youth aged 4–15 years with a reported diagnosis of hyperkinetic disorder, diagnosed for the first time in 1995–2005. 4–15 years.	372 (cases only)	Health administrative database[a]: psychiatric hospital data	1995–2005	ICD-10 F90.x
Morkem, 2020[35] Canada	Children and adults aged 4 and older identified from a single clinic, with a valid entry for year of birth and gender, and a primary care encounter in the year of study or previous year (from 2008–2015). Patients with certain medical conditions were excluded. ≥ 4 years.	246 (cases) 246 (non-cases)	Clinical/health information system[c]: electronic health records	NR	ICD-9 314.x Prescriptions of ADHD-related medications

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; EMR, electronic medical record; ICD-8, International Classification of Diseases, Eighth Revision; ICD-9, International Classification of Diseases, Ninth Revision; ICD-9-CM, International Classification of Diseases, Ninth Revision, clinical modification; ICD-10, International Classification of Diseases, Tenth Revision; ICD-10-CA, International Classification of Diseases, Tenth Revision, Canada; NR, not reported; OHIP, Ontario Health Insurance Plan.

[a] Health administrative database is defined as information passively collected, often by government and health care providers, for the purpose of managing the health care of patients (e.g. claims data).

[b] This study also tested algorithms that included case identification information from a keyword search of the cumulative patient profile in the electronic health record; however, these algorithms were not included, as they did not meet the review's definition of an administrative database algorithm.

[c] Clinical/health information system is defined as administrative data incorporating electronic health records, or, administrative data supplemented with detailed clinical information.

test results by disease severity, just three reported at least four measures of diagnostic accuracy,[22-24] only four reported the diagnostic accuracy by subgroup of interest[25-28] and just two of nine that reported the PPV and/or NPV reported a ratio of cases to controls in the validation cohort that approximates the prevalence of ASD in the population.[23,24]

ADHD studies

With respect to the methods used, none of the ADHD studies described the severity of the patients, none re-validated the algorithms using a separate cohort and only one reported that the reviewers of the reference standard were blinded to the administrative data classification.[35] Concerning the results, none reported test results by disease severity, just one

reported at least four measures of diagnostic accuracy,[33] only one stated the diagnostic accuracy by subgroups of interest,[32] and none reported a ratio of cases to controls in the validation cohort that approximated the prevalence of ADHD in the population.

Risk of bias and applicability concerns of included studies

An overview of the risk of bias and applicability concerns of the included studies by QUADAS-2 domain is shown in Figure 2.[20] Assessments revealed either "high" or "unclear" risk of bias in patient selection, reference standard and flow and timing domains in 5 or more of the 14 studies. All studies had a low risk of bias on the index test domain because of the objectivity of

administrative database algorithms. There were no applicability concerns with respect to the patient selection, index test or reference standard differing from the review question. For complete risk of bias and applicability assessments for each included study, see Appendix C.

ASD studies

Patient selection

Three studies had a high risk of bias,[22,25,31] one had a high risk in one of the two samples within the study,[27] and one had an unclear risk.[26] These evaluations were either due to the sampling approach,[25,27] insufficient information,[26] the use of a case-control design[22] or inappropriate exclusions.[31]

Reference standard

Two studies had an unclear risk of bias,[25,29] and one study had an unclear risk in one

Section, topic and item	Frequency (%)	
	ASD studies[b]	ADHD studies[c]
TITLE, KEYWORDS, ABSTRACT		
1. Identifies article as study of assessing diagnostic accuracy?	10 (100)	4 (100)
2. Identifies article as study of administrative data?	8 (80)	3 (75)
INTRODUCTION		
3. States disease identification and validation as one of goals of study?	10 (100)	4 (100)
METHODS		
Participants in validation cohort		
4. Describes validation cohort (cohort of patients to which reference standard was applied)?	10 (100)	4 (100)
4a. Age?	10 (100)	4 (100)
4b. Disease?	10 (100)	4 (100)
4c. Severity?	0 (0)	0 (0)
4d. Location/jurisdiction?	7 (70)	2 (50)
5. Describes recruitment procedure of validation cohort?	10 (100)	4 (100)
5a. Inclusion criteria?	10 (100)	4 (100)
5b. Exclusion criteria?	5 (50)	4 (100)
6. Describes patient sampling (random, consecutive, all, etc.)?	9 (90)	4 (100)
7. Describes data collection? (n = 8 ASD studies)	8 (100)	4 (100)
7a. Who identified patients and ensured selection adhered to patient recruitment criteria? (n = 8 ASD studies)	8 (100)	4 (100)
7b. Who collected data? (n = 8 ASD studies)	8 (100)	4 (100)
7c. A priori data collection form? (n = 8 ASD studies)	5 (62.5)	2 (50)
7d. How was disease classified?	10 (100)	3 (75)
8. Was there a split sample (i.e. re-validation using a separate cohort)?	1 (10)	0 (0)
Test methods		
9. Describe number, training and expertise of persons reading reference standard? (n = 8 ASD studies)	6 (75)	3 (75)
10. If > 1 person reading reference standard, measure of consistency is reported (e.g. kappa)? (n = 6 ASD studies; n = 3 ADHD studies)	3 (50)	2 (66.7)
11. Were the readers of the reference (validation) test blinded to the results of the classification by administrative data for that patient? (e.g. Was the reviewer of the charts blinded to how that chart was billed?) (n = 8 ASD studies)	3 (37.5)	1 (25)
Statistical methods		
12. Describe methods of calculating/comparing diagnostic accuracy?	10 (100)	3 (75)
RESULTS		
Participants		
13. Report when study done, start/end dates of enrolment?	8 (80)	2 (50)
14. Describe number of people who satisfied inclusion/exclusion criteria?	10 (100)	4 (100)
15. Study flow diagram?	4 (40)	3 (75)
Test results		
16. Report distribution of disease severity?	0 (0)	0 (0)
17. Report cross-tabulation of index tests by results of reference standard?	9 (90)	3 (75)

TABLE 2 (continued)

Number and percentage of included studies meeting individual modified STARD[a] reporting criteria for validating health administrative data

Section, topic and item	Frequency (%)	
	ASD studies[b]	ADHD studies[c]
Estimates		
18. Reports at least 4 estimates of diagnostic accuracy? (Estimates reported in included studies)	3 (30)	1 (25)
18a. Sensitivity	5 (50)	1 (25)
18b. Specificity	4 (40)	1 (25)
18c. PPV	9 (90)	4 (100)
18d. NPV	4 (40)	2 (50)
18e. Likelihood ratios	0 (0)	0 (0)
18f. Kappa	1 (10)	1 (25)
18g. Area under the ROC curve / C-statistic	2 (20)	0 (0)
18h. Accuracy/agreement	0 (0)	1 (25)
19. Was the accuracy reported for any subgroups (e.g. age, geography, different sex etc.)?	4 (40)	1 (25)
20. If PPV/NPV reported, does ratio of cases/controls of validation cohort approximate prevalence of condition in the population? (n = 9 ASD studies)	2 (22.2)	0 (0)
21. Reports 95% CIs for each diagnostic accuracy measure?	6 (60)	3 (75)
DISCUSSION		
22. Discusses the applicability of the findings?	10 (100)	4 (100)

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic.

Note: Modified STARD reporting criteria are from Benchimol et al.[12]

[a] Standards for the Reporting of Diagnostic Accuracy Studies (STARD): method of assessing reporting quality of validation studies using administrative data.

[b] n = 10 unless otherwise stated.

[c] n = 4 unless otherwise stated.

of its two samples.[27] These judgments were either due to insufficient information about the rigour of the reference standard,[27,29] a lack of information as to whether the reviewers were blinded to the results of the algorithm[25] or the reference standard being partly based on parent-reported diagnosis.[27]

Flow and timing

Five studies had a high risk of bias,[26-28,30,31] as not all patients were included in the analysis,[26,28,30,31] or not all patients were evaluated using the same reference standard.[27]

ADHD studies
Patient selection

Three studies had a high risk of bias due to inappropriate exclusions.[32,33,35]

Reference standard

One study had a high risk of bias, as the reference standard was not likely to classify cases correctly and reviewers were not blinded to the algorithm results,[33] and one study had an unclear risk of bias due to insufficient information.[35]

Flow and timing

Two studies had a high risk of bias,[33,34] as not all patients were included in the analysis[34] or not all patients were evaluated using the same reference standard.[33]

Diagnostic accuracy of administrative database algorithms

Given the heterogeneity found in study design and conduct across the included studies, the following synthesis highlights findings on the diagnostic accuracy of algorithms tested within, rather than between, studies. The diagnostic accuracy estimates of the algorithms varied substantially between studies, likely due to the observed between-study heterogeneity. Sources of this heterogeneity included differences in how cases were initially selected, administrative data sources, reference standards and algorithms tested. For example, two studies[23,24] with the same validation cohort and similar algorithms used different administrative data sources (health administrative database vs. clinical/health information system), observed very different performance

metrics, namely sensitivity and PPV. For the diagnostic accuracy of the algorithms validated in each included study, refer to Table 3.

ASD studies

For studies on ASD, the diagnostic accuracy of the algorithms tested was summarized in three different ways.

By health administrative database algorithm

Seven studies tested and compared multiple algorithms, each requiring more or fewer diagnostic codes from a specific health administrative data source (i.e. physician claims) over a comparable time frame.[22-28] In general, these studies found that increasing the number of ASD diagnoses required from physician claims increased the specificity and PPV of the algorithm, at the expense of sensitivity. For example, one study found a sensitivity of 62.5% and specificity of 83.0% when using an algorithm that required at least one ASD code from either the hospital or physician claims database.[28] However, when the same algorithm required at least two ASD codes from the physician claims

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
ASD							
Bickford, 2020[22]	☹	☺	☺	☺	☺	☺	☺
Brooks, 2021[23]	☺	☺	☺	☺	☺	☺	☺
Brooks, 2021[24]	☺	☺	☺	☺	☺	☺	☺
Burke, 2014[25]	☹	☺	?	☺	☺	☺	☺
Coleman, 2015[26]	?	☺	☺	☹	☺	☺	☺
Coo, 2017[27]	☹[b] ☺[c]	☺	?[b] ☹[c]	☹	☺	☺	☺
Dodds, 2009[28]	☺	☺	☺	☹	☺	☺	☺
Hagberg, 2017[29]	☺	☺	?	☺	☺	☺	☺
Lauritsen, 2010[30]	☺	☺	☺	☹	☺	☺	☺
Surén, 2019[31]	☹	☺	☺	☹	☺	☺	☺
ADHD							
Daley, 2014[32]	☹	☺	☺	☺	☺	☺	☺
Gruschow, 2016[33]	☹	☺	☹	☹	☺	☺	☺
Mohr-Jensen, 2016[34]	☺	☺	☺	☹	☺	☺	☺
Morkem, 2020[35]	☹	☺	?	☺	☺	☺	☺

☺ Low risk   ☹ High risk   ? Unclear risk

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder.

[a] Revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool: method of assessing the risk of bias and applicability of diagnostic studies through four key study domains, with each domain rated as having low, high or unclear risk of bias, and with low, high or unclear applicability to the research question.

[b] The risk of bias for the portion of the study evaluating sensitivity using the "sensitivity cohort" from the study.

[c] The risk of bias for the portion of the study evaluating positive predictive value (PPV) using children with an administrative diagnosis of ASD.

database, the specificity improved (93.2%) at the cost of a dramatic reduction in sensitivity (36.9%).

Three studies tested the value of additional health administrative data sources in their algorithms.[22,23,28] Two of these studies did not find a significant improvement in the diagnostic accuracy of those algorithms that required diagnostic codes from a combination of hospital discharge abstracts or physician claims (with or without emergency department visits or outpatient surgery) compared to physician claims alone.[22,23] One study that required at least one diagnostic code from one of three data sources (hospital discharge abstracts, mental health outpatient data or physician claims) increased the sensitivity of the algorithm by 9.6%, at the expense of specificity (7.9% decrease), compared to physician claims only.[28] Additionally, upon testing the accuracy of algorithms based on these three data sources separately (i.e. physician claims only, hospital discharge abstract only, mental health outpatient data only), the same study found that ASD diagnostic codes from physician claims led to the best-performing algorithm.

Two studies varied the number of years in which ASD diagnostic codes from physician claims were required in their algorithms (e.g. two or more codes in two years vs. two or more codes in three years).[23,24] Both of these studies found that increasing the number of years in which the codes could be found did not result in significant improvement in the diagnostic accuracy.

By reference standard

Of the 10 included studies, two varied the diagnostic criteria required for ASD case confirmation from more to less stringent.[25,26] Both of these studies found that when the evidence of ASD required in the medical chart was less stringent, the PPV increased substantially. For example, one study increased the PPV from 27% to 72% for an algorithm requiring at least one ASD code, and from 36% to 87% for an algorithm requiring at least two ASD codes.[26]

By combining education and health administrative data

Only one study validated algorithms using education and health administrative data to identify ASD.[27] In general, the algorithms that combined education and

TABLE 3
Diagnostic accuracy results of included studies

First author, year Country	Reference standard	Administrative database algorithm(s)	Measures of diagnostic accuracy (95% CI)[a]
ASD			
Bickford, 2020[22] Canada	Clinical diagnosis: Clinical data on ASD status from either the British Columbia Autism Assessment Network or the Ministry of Education. All diagnoses made using a standard approach based on DSM criteria, utilizing direct assessment of the child, information provided by family, and any other relevant information. Diagnoses were made by clinicians using the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview–Revised (ADI-R).	≥ 1 hospital discharge or ≥ 1 physician claim	SENS: 75% (74%–76%), SPEC: 67% (65%–69%), PPV: 82.7%, NPV: 55.6%, C-stat: 0.71 (0.70–0.72), Kappa: 0.40 (0.38–0.42)
		≥ 1 physician claim	SENS: 74% (72%–75%), SPEC: 68% (66%–69%), PPV: 82.7%, NPV: 54.7%, C-stat: 0.71 (0.70–0.72), Kappa: 0.39 (0.37–0.41)
		≥ 1 hospital discharge or ≥ 2 general practitioner claims or ≥ 1 pediatrician claim or ≥ 1 psychiatrist/neurologist claim or ≥ 1 other specialist claim	SENS: 67% (66%–69%), SPEC: 71% (69%–73%), PPV: 83.1%, NPV: 50.8%, C-stat: 0.69 (0.68–0.70), Kappa: 0.35 (0.33–0.37)
		≥ 1 hospital discharge or ≥ 3 general practitioner claims or ≥ 1 pediatrician claim or ≥ 1 psychiatrist/neurologist claim or ≥ 1 other specialist claim	SENS: 64% (63%–66%), SPEC: 73% (71%–74%), PPV: 83.2%, NPV: 49.1%, C-stat: 0.68 (0.67–0.69), Kappa: 0.33 (0.31–0.35)
		≥ 1 hospital discharge or ≥ 2 physician claims	SENS: 57% (56%–58%), SPEC: 84% (83%–86%), PPV: 88.4%, NPV: 48.2%, C-stat: 0.71 (0.70–0.72), Kappa: 0.35 (0.33–0.36)
		≥ 2 physician claims	SENS: 55% (54%–57%), SPEC: 85% (83%–86%), PPV: 88.5%, NPV: 47.3%, C-stat: 0.70 (0.69–0.71), Kappa: 0.33 (0.32–0.35)
		≥ 1 pediatrician claim	SENS: 54% (53%–56%), SPEC: 76% (75%–78%), PPV: 82.8%, NPV: 44.2%, C-stat: 0.65 (0.64–0.66), Kappa: 0.26 (0.24–0.28)
		≥ 1 hospital discharge or ≥ 2 general practitioner claims or ≥ 2 pediatrician claims or ≥ 2 psychiatrist/neurologist claims or ≥ 2 other specialist claims	SENS: 52% (51%–54%), SPEC: 86% (85%–87%), PPV: 88.7%, NPV: 46.1%, C-stat: 0.69 (0.68–0.70), Kappa: 0.31 (0.30–0.33)
		≥ 2 general practitioner claims or ≥ 2 pediatrician claims or ≥ 2 psychiatrist/neurologist claims or ≥ 2 other specialist claims	SENS: 50% (49%–52%), SPEC: 87% (85%–88%), PPV: 88.8%, NPV: 45.2%, C-stat: 0.68 (0.68–0.69), Kappa: 0.30 (0.28–0.31)
		≥ 1 general practitioner claim	SENS: 44% (42%–45%), SPEC: 89% (88%–90%), PPV: 89.5%, NPV: 42.8%, C-stat: 0.66 (0.66–0.67), Kappa: 0.26 (0.24–0.27)
		≥ 1 psychiatrist/neurologist claim	SENS: 14% (13%–15%), SPEC: 97% (96%–97%), PPV: 89.6%, NPV: 34.8%, C-stat: 0.55 (0.55–0.56), Kappa: 0.07 (0.07–0.08)
Brooks, 2021[23] Canada	Medical chart review—ASD diagnosis: Manual review of electronic medical record for diagnosis of ASD. Cases were identified by a trained nurse chart abstractor and confirmed by a family physician.	≥ 1 hospital discharge or ≥ 1 emergency department visit or ≥ 1 outpatient surgery or ≥ 1 physician claim	SENS: 75.9% (68.0%–83.8%), SPEC: 98.9% (98.6%–99.1%), PPV: 42.9% (36.0%–49.8%), NPV: 99.7% (99.6%–99.8%)
		≥ 1 physician claim	SENS: 74.1% (66.0%–82.2%), SPEC: 98.9% (98.7%–99.1%), PPV: 42.6% (35.6%–49.5%), NPV: 99.7% (99.6%–99.8%)
		≥ 1 hospital discharge or ≥ 1 emergency department visit or ≥ 1 outpatient surgery or ≥ 1 physician claim by any specialist	SENS: 67.9% (59.2%–76.5%), SPEC: 99.0% (98.9%–99.2%), PPV: 44.7% (37.2%–52.2%), NPV: 99.6% (99.5%–99.8%)
		≥ 1 physician claim by any specialist	SENS: 66.1% (57.3%–74.8%), SPEC: 99.1% (98.9%–99.2%), PPV: 44.3% (36.8%–51.8%), NPV: 99.6% (99.5%–99.7%)
		≥ 1 hospital discharge or ≥ 1 emergency department visit or ≥ 1 outpatient surgery; or ≥ 2 physician claims in 3 years	SENS: 59.8% (50.7%–68.9%), SPEC: 99.3% (99.1%–99.5%), PPV: 49.3% (40.9%–57.7%), NPV: 99.5% (99.4%–99.7%)

Continued on the following page

TABLE 3 (continued)
Diagnostic accuracy results of included studies

First author, year Country	Reference standard	Administrative database algorithm(s)	Measures of diagnostic accuracy (95% CI)[a]
		≥ 1 hospital discharge or ≥ 1 emergency department visit or ≥ 1 outpatient surgery; or ≥ 2 physician claims in 2 years	SENS: 57.1% (48.0%–66.3%), SPEC: 99.3% (99.1%–99.5%), PPV: 48.5% (40.0%–57.0%), NPV: 99.5% (99.4%–99.7%)
		≥ 1 hospital discharge or ≥ 1 emergency department visit or ≥ 1 outpatient surgery; or ≥ 2 physician claims in 3 years with ≥ 1 from any specialist	SENS: 53.6% (44.3%–62.8%), SPEC: 99.4% (99.2%–99.5%), PPV: 48.4% (39.6%–57.2%), NPV: 99.5% (99.3%–99.6%)
		≥ 1 hospital discharge or ≥ 1 emergency department visit or ≥ 1 outpatient surgery; or ≥ 2 physician claims in 2 years with ≥ 1 from any specialist	SENS: 52.7% (43.4%–61.9%), SPEC: 99.4% (99.2%–99.5%), PPV: 48.4% (39.5%–57.2%), NPV: 99.5% (99.3%–99.6%)
		≥ 1 hospital discharge or ≥ 1 emergency department visit or ≥ 1 outpatient surgery; or ≥ 3 physician claims in 3 years	SENS: 50.0% (40.7%–59.3%), SPEC: 99.6% (99.4%–99.7%), PPV: 56.6% (46.8%–66.3%), NPV: 99.4% (99.3%–99.6%)
		≥ 1 hospital discharge or ≥ 1 emergency department visit or ≥ 1 outpatient surgery; or ≥ 3 physician claims in 3 years with ≥ 1 from any specialist	SENS: 49.1% (39.8%–58.4%), SPEC: 99.6% (99.5%–99.7%), PPV: 57.9% (48.0%–67.8%), NPV: 99.4% (99.3%–99.6%)
		≥ 1 hospital discharge or ≥ 1 emergency department visit or ≥ 1 outpatient surgery; or ≥ 3 physician claims in 2 years	SENS: 45.5% (36.3%–54.8%), SPEC: 99.6% (99.4%–99.7%), PPV: 54.3% (44.2%–64.3%), NPV: 99.4% (99.2%–99.5%)
		≥ 1 hospital discharge or ≥ 1 emergency department visit or ≥ 1 outpatient surgery; or ≥ 3 physician claims in 2 years with ≥ 1 from any specialist	SENS: 45.5% (36.3%–54.8%), SPEC: 99.6% (99.5%–99.7%), PPV: 56.0% (45.8%–66.2%), NPV: 99.4% (99.2%–99.5%)
Brooks, 2021[24,b] Canada	Medical chart review—ASD diagnosis: Manual review of electronic medical record for diagnosis of ASD. Cases were identified by a trained nurse chart abstractor and confirmed by a family physician.	≥ 1 physician claim (299 or 315)	SENS: 33.0% (24.4%–42.6%), SPEC: 98.8% (98.5%–99.0%), PPV: 23.4% (17.1%–30.8%), NPV: 99.2% (99.0%–99.4%)
		≥ 2 physician claims (299 or 315) in 3 years	SENS: 14.3% (8.4%–22.2%), SPEC: 99.8% (99.7%–99.9%), PPV: 44.4% (27.9%–61.9%), NPV: 99.0% (98.8%–99.2%)
		≥ 2 physician claims (299 or 315) in 2 years	SENS: 13.4% (7.7%–21.1%), SPEC: 99.8% (99.7%–99.9%), PPV: 45.5% (28.1%–63.6%), NPV: 99.0% (98.8%–99.2%)
		≥ 2 physician claims (299 or 315) in 1 year	SENS: 11.6% (6.3%–19.0%), SPEC: 99.9% (99.8%–99.9%), PPV: 52.0% (31.3%–72.2%), NPV: 99.0% (98.8%–99.2%)
		≥ 3 physician claims (299 or 315) in 3 years	SENS: 2.7% (0.6%–7.6%), SPEC: 99.9% (99.9%–100%), PPV: 33.3% (7.5%–70.1%), NPV: 98.9% (98.7%–99.1%)
		≥ 3 physician claims (299 or 315) in 2 years	SENS: 2.7% (0.6%–7.6%), SPEC: 99.9% (99.9%–100%), PPV: 33.3% (7.5%–70.1%), NPV: 98.9% (98.7%–99.1%)
		≥ 3 physician claims (299 or 315) in 1 year	SENS: 1.8% (0.2%–6.3%), SPEC: 100% (99.9%–100%), PPV: 33.3% (4.3%–77.7%), NPV: 98.9% (98.7%–99.1%)
		≥ 1 physician claim (299 only)	SENS: 28.6% (20.4%–37.9%), SPEC: 99.9% (99.9%–100%), PPV: 86.5% (71.2%–95.5%), NPV: 99.2% (99.0%–99.4%)
		≥ 2 physician claims (299 only) in 3 years	SENS: 12.5% (7.0%–20.1%), SPEC: 100% (99.9%–100%), PPV: 93.3% (68.1%–99.8%), NPV: 99.0% (98.8%–99.2%)

Continued on the following page

TABLE 3 (continued)
Diagnostic accuracy results of included studies

First author, year Country	Reference standard	Administrative database algorithm(s)	Measures of diagnostic accuracy (95% CI)[a]
		≥ 2 physician claims (299 only) in 2 years	SENS: 11.6% (6.3%–19.0%), SPEC: 100% (99.9%–100%), PPV: 92.9% (66.1%–99.8%), NPV: 99.0% (98.8%–99.2%)
		≥ 2 physician claims (299 only) in 1 year	SENS: 9.8% (5.0%–16.9%), SPEC: 100% (99.9%–100%), PPV: 91.7% (61.5%–99.8%), NPV: 99.0% (98.8%–99.2%)
		≥ 3 physician claims (299 only) in 3 years	SENS: 1.8% (0.2%–6.3%), SPEC: 100% (99.9%–100%), PPV: 66.7% (9.4%–99.2%), NPV: 98.9% (98.7%–99.1%)
		≥ 3 physician claims (299 only) in 2 years	SENS: 1.8% (0.2%–6.3%), SPEC: 100% (99.9%–100%), PPV: 66.7% (9.4%–99.2%), NPV: 98.9% (98.7%–99.1%)
		≥ 3 physician claims (299 only) in 1 year	SENS: 0.9% (0.0%–4.9%), SPEC: 100% (100%–100%), PPV: 100% (2.5%–100%), NPV: 98.9% (98.7%–99.1%)
Burke, 2014[25] USA	Medical chart review—clinical classification criteria, ASD diagnosis: Criteria used to confirm ASD: (1) level 1—behavioural descriptions highly indicative of ASD and consistent with DSM-IV-TR criteria; or (2) level 2—provider documented diagnosis or some evidence of ASD behaviours consistent with DSM-IV-TR criteria (but not enough description to qualify as level 1).	≥ 1 ASD-associated condition (no ASD insurance claim)	NPV (level 1 or 2): > 98%
		≥ 1 ASD insurance claim	PPV (level 1): 43.3% (38.2%–48.5%) PPV (level 1 or 2): 74.2% (69.4%–78.6%)
		≥ 2 ASD insurance claims	PPV (level 1): 60.9% (53.5%–68.1%) PPV (level 1 or 2): 87.4% (81.6%–91.8%)
Coleman, 2015[26] USA	Medical chart review—clinical classification criteria, ASD diagnosis: Criteria used to confirm ASD: (1) confirmed—complete, documented assessment using DSM-IV criteria; (2) probable—diagnosis made by a credible source, documented use of DSM-IV to make diagnosis, and some documented patient behaviours consistent with DSM-IV criteria; or (3) possible—second-hand reports of an ASD assessment by a professional or some documented behaviours associated with ASD.	1 insurance claim or outpatient diagnosis	PPV[c] (confirmed): 27% PPV[c] (confirmed, probable and possible): 72%
		≥ 2 insurance claims or outpatient diagnoses, at least one day apart	PPV[c] (confirmed): 36% PPV[c] (confirmed, probable and possible): 87%
Coo, 2017[27] Canada	Medical chart review—ASD diagnosis: Review of chart/file for an ASD diagnosis by one of four child/youth behavioural or disability service providers. Parent-reported diagnosis: For children with an administrative diagnosis of ASD but no confirmed diagnosis on chart/file, parent-reported diagnoses were also considered true positives.	Ages 2–5:	Ages 2–5:
		≥ 1 physician claim or ≥ 1 "ASD" education code or ≥ 1 hospital discharge or ≥ 1 adolescent treatment centre diagnosis	SENS: 88% (84%–91%), minimum PPV[d]: 73% (68%–77%)
		≥ 1 physician claim or ≥ 1 "ASD" education code	SENS: 88% (83%–91%), minimum PPV[d]: 73% (69%–78%)
		≥ 1 physician claim	SENS: 85% (80%–88%), minimum PPV[d]: 73% (68%–77%)
		≥ 2 physician claims or ≥ 1 "ASD" education code	SENS: 57% (51%–62%), minimum PPV[d]: 89% (84%–93%)
		≥ 2 physician claims	SENS: 50% (45%–56%), minimum PPV[d]: 89% (83%–93%)

TABLE 3 (continued)
Diagnostic accuracy results of included studies

First author, year Country	Reference standard	Administrative database algorithm(s)	Measures of diagnostic accuracy (95% CI)[a]
		Ages 6–9:	**Ages 6–9:**
		≥ 1 physician claim or ≥ 1 "ASD" education code or ≥ 1 hospital discharge or ≥ 1 adolescent treatment centre diagnosis	SENS: 90% (88%–93%), minimum PPV[d]: 65% (61%–68%)
		≥ 1 physician claim or ≥ 1 "ASD" education code	SENS: 89% (86%–92%), minimum PPV[d]: 65% (61%–68%)
		≥ 1 physician claim or ≥ 2 "ASD" education codes	SENS: 88% (85%–90%), minimum PPV[d]: 65% (62%–69%)
		≥ 2 physician claims or ≥ 1 "ASD" education code	SENS: 84% (81%–87%), minimum PPV[d]: 78% (75%–81%)
		≥ 2 physician claims or ≥ 2 "ASD" education codes	SENS: 81% (78%–84%), minimum PPV[d]: 80% (77%–84%)
		≥ 1 physician claim	SENS: 77% (73%–80%), minimum PPV[d]: 66% (62%–69%)
		≥ 1 "ASD" education code	SENS: 68% (64%–72%), minimum PPV[d]: 87% (84%–90%)
		≥ 2 "ASD" education codes	SENS: 66% (62%–70%), minimum PPV[d]: 88% (85%–91%)
		≥ 2 physician claims	SENS: 58% (54%–62%), minimum PPV[d]: 83% (79%–86%)
		Ages 10–14:	**Ages 10–14:**
		≥ 1 physician claim or ≥ 1 "ASD" education code or ≥ 1 hospital discharge or ≥ 1 adolescent treatment centre diagnosis	SENS: 88% (85%–90%), minimum PPV[d]: 60% (57%–63%)
		≥ 1 physician claim or ≥ 1 "ASD" education code	SENS: 86% (83%–88%), minimum PPV[d]: 61% (58%–63%)
		≥ 1 physician claim or ≥ 2 "ASD" education codes	SENS: 84% (82%–87%), minimum PPV[d]: 62% (59%–64%)
		≥ 2 physician claims or ≥ 1 "ASD"' education code	SENS: 80% (77%–83%), minimum PPV[d]: 70% (67%–73%)
		≥ 2 physician claims or ≥ 2 "ASD" education codes	SENS: 78% (75%–81%), minimum PPV[d]: 72% (69%–75%)
		≥ 1 physician claim	SENS: 73% (69%–76%), minimum PPV[d]: 64% (60%–70%)
		≥ 1 "ASD" education code	SENS: 73% (70%–76%), minimum PPV[d]: 75% (72%–78%)
		≥ 2 "ASD" education codes	SENS: 69% (66%–72%), minimum PPV[d]: 78% (74%–81%)
		≥ 2 physician claims	SENS: 56% (52%–59%), minimum PPV[d]: 78% (75%–82%)
Dodds, 2009[28] Canada	Clinical diagnosis: Clinical diagnosis by a team of ASD specialists; based on the Autism Diagnostic Interview–Revised, the Autism Diagnostic Observation Schedule and clinical judgment using DSM-IV-TR.	≥ 1 hospital discharge or ≥ 1 physician claim or ≥ 1 mental health outpatient diagnosis	SENS: 69.3%, SPEC: 77.3%, C-stat: 0.76
		≥ 1 hospital discharge or ≥ 1 physician claim	SENS: 62.5%, SPEC: 83.0%, C-stat: 0.74
		≥ 1 physician claim	SENS: 59.7%, SPEC: 85.2%, C-stat: 0.72
		≥ 1 hospital discharge or ≥ 2 physician claims or ≥ 2 mental health outpatient diagnoses	SENS: 42.6%, SPEC: 88.6%, C-stat: 0.67
		≥ 1 hospital discharge or ≥ 2 physician claims	SENS: 36.9%, SPEC: 93.2%, C-stat: 0.65
		≥ 1 mental health outpatient diagnosis	SENS: 16.5%, SPEC: 92.0%, C-stat: 0.54
		≥ 1 hospital discharge diagnosis	SENS: 11.9%, SPEC: 97.7%, C-stat: 0.55

TABLE 3 (continued)
Diagnostic accuracy results of included studies

First author, year Country	Reference standard	Administrative database algorithm(s)	Measures of diagnostic accuracy (95% CI)[a]
Hagberg, 2017[29] United Kingdom	Medical chart review—ASD diagnosis: Review of original medical record for diagnosis confirmation, which included detailed hospital clinical letters, consultant reports, speech and language assessments and/or specialist reports.	≥ 1 Read code	PPV: 91.9%
Lauritsen, 2010[30] Denmark	Medical chart review—clinical classification criteria: Modified version of the CDC coding guide based on DSM-IV, i.e. scored positive on at least one social and either one communication or one behavioural criterion, with no diagnoses, history or behavioural descriptions that contradicted the presence of ASD.	1 psychiatric inpatient or outpatient diagnosis	PPV: 97% (96%–99%)
Surén, 2019[31] Norway	Medical chart review—clinical classification criteria: Review of patient records to confirm if child met ICD-10 diagnostic criteria; included results of standardized interviews/tests and diagnoses received.	≥ 1 diagnosis	PPV: 86% (83%–89%)
ADHD			
Daley, 2014[32] USA	Medical chart review—clinical classification criteria, ADHD diagnosis, standardized screening checklist: Criteria used to confirm ADHD: (1) definition 1—clinician diagnosis in either the index or follow-up window; (2) definition 2—clinician diagnosis in either the index or follow-up window, with prevalent cases excluded; (3) definition 3—clinician diagnosis, prevalent cases excluded, at least one positive ADHD screening checklist; (4) definition 4—clinician diagnosis, prevalent cases excluded, at least 6 of 9 inattentive and/or 6 of 9 hyperactive/impulsive DSM-IV symptoms; or (5) definition 5—clinician diagnosis, prevalent cases excluded, at least one positive screening checklist or at least 6 of 9 inattentive and/or 6 of 9 hyperactive/impulsive DSM-IV symptoms.	2 outpatient diagnoses, between 7 and 365 days apart (incident ADHD)	Ages 3–5 at diagnosis: PPV[e] (definition 1): 89.8% (80.6%–99.0%) PPV[e] (definition 2): 71.5% (56.5%–86.4%) PPV[e] (definition 3): 48.9% (33.4%–64.3%) PPV[e] (definition 4): 32.8% (17.1%–48.5%) PPV[e] (definition 5): 65.8% (52.2%–79.4%) Ages 6–9 at diagnosis: PPV[e] (definition 1): 94.2% (89.8%–98.5%) PPV[e] (definition 2): 73.6% (65.6%–81.6%) PPV[e] (definition 3): 59.1% (50.8%–67.5%) PPV[e] (definition 4): 30.9% (22.2%–39.6%) PPV[e] (definition 5): 68.5% (60.8%–76.1%)
Gruschow, 2016[33] USA	Clinical case definition:[f] Patients with ADHD confirmed if: ≥ 3 ADHD-related visits, or 1 or 2 ADHD-related visits or a problem list diagnosis and prescribed ADHD medication, or evidence from an independent source confirming ADHD case status located through a manual review of the electronic health record. Patients without ADHD confirmed when evidence from an independent source indicated the patient did not have ADHD through a manual review of the electronic health record.	≥ 1 inpatient or outpatient diagnosis or problem list[g] diagnosis	SENS: 96%–97% (95%–97%), SPEC: 98%–99% (97%–99%), PPV: 83%–98% (81%–99%), NPV: 99% (99%–99%),[h] Kappa: 0.87 (0.75–0.99)

Continued on the following page

TABLE 3 (continued)
Diagnostic accuracy results of included studies

First author, year Country	Reference standard	Administrative database algorithm(s)	Measures of diagnostic accuracy (95% CI)[a]
Mohr-Jensen, 2016[33] Denmark	Medical chart review—clinical classification criteria: Patient files were systematically scored for the presence of ICD-10 criteria for hyperkinetic disorder and were confirmed if patients presented with ≥ 6 symptoms of inattention, ≥ 3 symptoms of hyperactivity and ≥ 1 symptom of impulsivity.	1 psychiatric inpatient or outpatient diagnosis	PPV: 86.8%
Morkem, 2020[35] Canada	Medical chart review—ADHD diagnosis: Review of electronic medical record for diagnosis of ADHD.	≥ 1 medical visit (ICD code) and ≥ 1 prescription of ADHD-related medications or ≥ 2 medical visits (ICD code)	PPV: 95.9% (92.6%–98.0%), NPV: 96.3% (93.2%–98.3%)

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; CDC, Centers for Disease Control and Prevention; CI, confidence interval; C-stat, C-statistic; DSM, *Diagnostic and Statistical Manual of Mental Disorders*; DSM-IV, *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Revision; DSM-IV-TR, *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Revision, Text Revision; ICD-10, International Classification of Diseases, Tenth Revision; ICD, International Classification of Diseases; Kappa, kappa statistic; NPV, negative predictive value; PPV, positive predictive value; SENS, sensitivity; SPEC, specificity.

[a] Where applicable.

[b] This study also tested algorithms that included case identification information from a keyword search of the cumulative patient profile in the electronic health record; however, these algorithms were not included, as they did not meet the review's definition of an administrative database algorithm.

[c] Cases without enough information were excluded from PPV calculations.

[d] Those for whom case status could not be ascertained were assumed to be false positives for the purpose of this calculation.

[e] All sampled cases were weighted by their inverse selection probability.

[f] Clinical case definition was based on a medical chart review or electronic health record and was not as stringent as clinical classification criteria.

[g] List of ongoing or historical problems in the patient's electronic health record.

[h] Range depends on assumptions for inconclusive cases, i.e. true ADHD cases vs. true non-ADHD cases.

physician claims data demonstrated an improvement in sensitivity, but the PPV either remained the same or decreased slightly compared to algorithms based on physician claims data alone. For example, in the group aged 6 to 9 years, requiring at least one code from physician claims or education data versus at least one physician code only caused a substantial increase in sensitivity (from 77% to 89%) and a nominal decrease in PPV (from 66% to 65%). A similar pattern was observed in the oldest age group, those aged 10 to 14 years.

ADHD studies

For studies on ADHD, the diagnostic accuracy of the algorithms tested was summarized in the same ways; however, none of the studies on ADHD used education data in addition to health administrative data, therefore any benefit of this data cannot be assessed.

By health administrative database algorithm

All four studies tested and reported results for one algorithm only and each of these algorithms included diagnostic codes from one administrative data source only.[32-35] As a result, there was not enough information to assess the impact of requiring more or fewer diagnostic codes or utilizing additional data sources in identifying ADHD cases.

By reference standard

One of the four studies varied the diagnostic criteria required for the reference standard and presented results for more and less stringent requirements for incident ADHD case confirmation.[32] As more documented evidence of incident ADHD was required, the PPV decreased from 71.5% to 32.8% in children aged 3 to 5 years at diagnosis and from 73.6% to 30.9% in children aged 6 to 9.

Discussion

A total of 14 studies met our eligibility criteria,[22-35] of which 10 focussed on the validation of administrative database algorithms to identify ASD[22-31] and four on ADHD.[32-35] Six of the 14 studies were conducted in Canada and had generally a higher reporting quality and lower risk of bias compared to studies from other countries.[22-24,27,28,35] There were no studies identified for FASD that met the eligibility criteria for our review. Other important gaps identified included a lack of validation studies on adults, and the identification of incident, rather than prevalent, cases.

While there have been efforts to use health administrative data to estimate the prevalence of FASD in Canada,[36] this work has been done in the absence of any validated administrative database algorithms. The lack of published validation studies for FASD may be connected to several fundamental issues related to assigning an FASD diagnosis, including:

- the need for a multidisciplinary assessment and knowledge of prenatal alcohol exposure;[37,38]

- the lack of diagnostic criteria or detailed description for the diagnosis "neurodevelopmental disorder associated with prenatal alcohol exposure" in the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition;[1,39] and

- the non-specific nature of International Classification of Diseases diagnostic codes that can be used for a primary FASD diagnosis,[39] and the specificity of codes required to capture diagnostic entities associated with FASD, which is not always available within health administrative databases.

Given the evolving nature of FASD diagnostic practices and coding, the use of administrative database algorithms to identify FASD cases currently poses some unique challenges.

Results from the quality assessments revealed "high" or "unclear" risk of bias in at least one domain of the QUADAS-2 tool for 12 of the 14 studies,[22,25-35] indicating the measures of diagnostic accuracy should be interpreted with caution. Furthermore, significant heterogeneity in terms of study design and conduct across the included studies prohibited a quantitative synthesis of the results on the accuracy of the algorithms tested. Of particular importance were the differences in how cases were initially selected (i.e. either by diagnostic codes in the administrative database or the reference standard) and the inclusion or exclusion of cases without the condition of interest. To ensure unbiased estimates of diagnostic accuracy, the disease prevalence in the validation cohort must approximate the prevalence in the population.[40] This is achieved when an appropriate reference standard is applied (which accurately classifies cases with and without the condition of interest) and patients are randomly sampled, ideally from the general population. In addition, to compute the key diagnostic accuracy estimates necessary to evaluate the characteristics of a diagnostic test, both cases with and without the condition of interest are needed to populate all four cells of a two-way contingency table.

Unfortunately, most studies in our review (8 out of 14) used diagnostic codes in the administrative database to initially select cases with the condition of interest only.[26,27,29-32,34,35] This approach can generate biased estimates of diagnostic accuracy, given the underlying prevalence of the conditions of interest is unknown, and it limits the diagnostic accuracy measures that can be computed to PPV only.[40] Although PPV defines the likelihood of false-positive test results, alone it does not provide information on the likelihood of false-negative test results or how many cases are being missed by the algorithm.

Furthermore, while seven of the 14 studies included cases without the condition of interest,[22-25,28,33,35] two drew their samples from specialty clinics or service providers,[22,28] and two oversampled children more likely to have the disorder;[25,33] both of these sampling methods can generate falsely elevated PPV due to the high prevalence of cases. In addition, only four of these studies reported the four key measures of diagnostic accuracy that can be computed using this approach.[22-24,33]

Despite these limitations, findings from our review suggest that increasing the number of ASD diagnostic codes required from physician claims database increases specificity and PPV of an algorithm, at the expense of sensitivity. In addition, the use of multiple sources of health administrative data in an algorithm designed to identify ASD cases (i.e. hospital, physician claims and mental health services) may increase sensitivity with only a slight cost to the specificity and PPV, with physician claims database being the best single source.

Furthermore, the findings from one study showed that the addition of education data, in combination with physician claims data, might improve case capture (sensitivity) in school-aged children and youth at a slight cost to precision (PPV).[27] However, the lack of cases without ASD in this study limited the diagnostic accuracy measures that could be computed. Therefore, additional studies are required to evaluate the full impact of including education data in combination with physician claims data in administrative database algorithms for ASD case ascertainment purposes.

Due to the nature of the ADHD studies included, there was not enough information to assess the impact of number of diagnostic codes or additional data sources on algorithm accuracy. However, based on the performance measures reported, there was some evidence that ADHD could be identified through health administrative data sources.

To address the gaps uncovered by this review, as well as the reporting quality and risk of bias issues found, additional high quality studies validating the use of administrative database algorithms to identify cases of the selected neurodevelopmental disorders are required. These issues are not unique to this area of study, and guidance on how to conduct and report the findings from such validation studies has been previously published.[12,20,40] In light of all of this, we recommend authors follow published recommendations on study methods[20,40] as well as reporting guidelines[12] when validating administrative database algorithms for case identification purposes.

Another challenge associated with the use of specific diagnostic codes for neurodevelopmental disorders such as ASD, ADHD and FASD relates to the fact that the boundaries between these disorders are often not clear and the presence of comorbid disorders is common. While historically neurodevelopmental disorders have been categorically diagnosed based on a constellation of signs and symptoms, there is an evolving body of literature on the need for new approaches to their diagnosis that involves conceptualizing these disorders as lying on a neurodevelopmental continuum.[41] This shift will have important implications on the classification of these disorders, clinical practice, research and surveillance.

Strengths and limitations

The strengths of this review include:

- its prospective registration with the Prospective Register of Systematic Reviews (PROSPERO), which helps to reduce the potential for bias in the conduct and reporting of systematic reviews;[42]

- the development of a literature search strategy by an experienced reference librarian that included a systematic search of multiple databases, the grey literature and reference lists of included articles;

- a rigorous assessment of the reporting quality as well as the risk of bias and applicability of each included study using the modified STARD checklist[12] and the QUADAS-2 tool,[20] respectively; and

- the use of the PRISMA standards to ensure full reporting and transparency.[17]

However, there are some limitations worth noting, such as:

- challenges in conducting a comprehensive search for studies focussing on administrative database algorithms, given they are not well catalogued in the databases we searched (i.e. no medical subject headings on "administrative database" exist);

- the potential for language bias, as studies published in a language other than English or French were not considered, as well as publication bias, since validation studies with poor results may be less likely to be published; and

- the significant heterogeneity between included studies did not permit the conduct of quantitative analyses such as a meta-regression or meta-analysis.

Conclusion

To our knowledge, this is the first review that has systematically appraised and examined the empirical evidence on validity of administrative database algorithms to identify ASD, ADHD and FASD. While a few studies have validated algorithms for ASD and ADHD case ascertainment purposes, none have been performed for FASD to date. Significant heterogeneity across included studies limited our ability to carry out quantitative analyses. Such analyses would be beneficial to further strengthen the evidence around the best-performing algorithms for neurodevelopmental disorders surveillance and research, should the quality of available studies allow.

Nevertheless, there is some evidence to suggest that ASD and ADHD can be identified using administrative data, although information about the ability to discriminate reliably between individuals with and without the disorder of interest is limited. Given the variations in reporting quality and risk of bias issues found, additional high quality validation studies are needed. To optimize the usefulness of future studies, we recommend authors follow published recommendations on study design and conduct[20,40] and reporting guidelines for validation studies involving administrative data.[12]

Conflicts of interest

None.

Authors' contributions and statement

CL and SO conceptualized and designed the study; SO and SP helped with developing the search strategy and retrieving articles; CL, ML and SO screened the literature; ML and SO were responsible for data extraction, reporting quality, risk of bias and applicability assessments; all authors analyzed and/or interpreted the data; SO and SP drafted the manuscript; and all authors contributed to the initial draft and revisions of the manuscript.

The content and views expressed in this article are those of the authors and do not necessarily reflect those of the Government of Canada.

INITIAL SEARCH

Database(s): **Ovid MEDLINE(R) ALL** 1946 to August 26, 2019
Search Strategy:

#	Searches	Results
1	exp autism spectrum disorder/ or (autism or autistic or (asperger* adj (syndrome or disorder or disease)) or kanner* syndrome or childhood disintegrative disorder or (pervasive adj2 developmental disorder*) or heller* syndrome or disintegrative psychosis).tw,kf,kw.	48476
2	Attention Deficit Disorder with Hyperactivity/ or "Attention Deficit and Disruptive Behavior Disorders"/ or ((attention deficit adj4 disorder*) or (hyperkinetic adj2 (disorder or syndrome)) or minimal brain dysfunction or adhd or addh).tw,kf,kw.	40083
3	Fetal Alcohol Spectrum Disorders/ or (f?etal alcohol or (alcohol related adj3 (neurodevelopment or birth)) or "Neurobehavioral disorder associated with prenatal alcohol exposure" or "Growth Retardation, Facial Abnormalities, and Central Nervous System Dysfunction").tw,kw,kf.	5699
4	or/1-3	90368
5	diagnosis/ or incidence/ or prevalence/ or (diagnos* or incidence* or prevalence* or new case?).tw,kw,kf.	3507353
6	4 and 5	27896
7	exp autism spectrum disorder/di, dg, ep or Attention Deficit Disorder with Hyperactivity/di, dg, ep or "Attention Deficit and Disruptive Behavior Disorders"/di, dg, ep or Fetal Alcohol Spectrum Disorders/di, dg, ep	22499
8	6 or 7 [Developmental disorders]	39288
9	exp medical records/ or international classification of diseases/ or exp diagnostic techniques, neurological/ or exp clinical laboratory techniques/ or "diagnostic techniques and procedures"/	3030824
10	(((patient* or medical or health) adj3 record*) or ((diagnos* or defin* or classificat*) adj2 disease list) or ((self report* or standard*) adj2 measure*) or (icd adj (cod* or classification*)) or (international adj3 classification)).tw,kw,kf.	248712
11	medicaid/ or birth certificates/ or death certificates/ or hospital records/ or insurance claim reporting/ or exp insurance, health/ or databases, factual/ or information systems/ or databases as topic/ or database management systems/ or software/ or insurance claim review/ or patient discharge/ or exp registries/ or utilization review/	472574
12	(((administrat* or physician* or inpatient* or emergency* or hospital* or clinic or clinics or pharmac* or insurance) adj4 (admission* or data or dataset* or database* or data base? or data bank? or claim* or billing* or record* or utilizat* or utilisat*)) or ((claim* or discharg*) adj2 data*) or ((database? or databank? or data base? or data bank?) adj4 (factual or administrat* or claim? or register* or registr* or topic? or system?)) or (claim? adj2 (analy* or review? or physician? or pharmac* or drug?)) or (insurance adj2 (claim* or audit*)) or medicaid or (health* adj2 plan) or ((death* or birth*) adj1 (certificate* or record*))).tw,kw,kf. or (database? or databank? or data base? or data bank?).ti.	290378
13	or/9-12 [Diagnosis Methods]	3827029
14	8 and 13	5662
15	exp Diagnostic Errors/ or Diagnosis, Differential/ or "Predictive Value of Tests"/ or "Sensitivity and Specificity"/ or ROC Curve/ or Area under Curve/ or Bayes Theorem/ or algorithms/ or validation studies as topic/	1272706
16	((diagnos* adj3 (schedul* or clinical* or technique* or procedur* or assess* or standard* or error* or false or incorrect* or wrong* or correct*)) or misdiagnos* or (clinical* adj (standard* or criteri* or measure* or classifi* or technique* or assess*)) or ((positive or negative) adj2 (predict* or false)) or sensitiv* or specif* or accura* or valid* or reliab* or agree* or concord* or misclass* or ((case or cases) adj2 ascertain*) or algorithm? or (bayes* adj (theorem or analysis or approach or forecast or method or prediction)) or (roc adj (curve or analysis)) or receiver operating characteristic).tw,kw,kf.	5708365
17	15 or 16	6369226
18	14 and 17	2476
19	limit 18 to (yr="1995-2019" and (english or french))	2196

FIRST SEARCH UPDATE

Database(s): **Ovid MEDLINE(R) ALL** 1946 to July 10, 2020
Search Strategy:

#	Searches	Results
1	exp autism spectrum disorder/ or (autism or autistic or (asperger* adj (syndrome or disorder or disease)) or kanner* syndrome or childhood disintegrative disorder or (pervasive adj2 developmental disorder*) or heller* syndrome or disintegrative psychosis).tw,kf,kw.	53121
2	Attention Deficit Disorder with Hyperactivity/ or "Attention Deficit and Disruptive Behavior Disorders"/ or ((attention deficit adj4 disorder*) or (hyperkinetic adj2 (disorder or syndrome)) or minimal brain dysfunction or adhd or addh).tw,kf,kw.	42517
3	Fetal Alcohol Spectrum Disorders/ or (f?etal alcohol or (alcohol related adj3 (neurodevelopment or birth)) or "Neurobehavioral disorder associated with prenatal alcohol exposure" or "Growth Retardation, Facial Abnormalities, and Central Nervous System Dysfunction"). tw,kw,kf.	5917
4	or/1-3	97183
5	diagnosis/ or incidence/ or prevalence/ or (diagnos* or incidence* or prevalence* or new case?).tw,kw,kf.	3713126
6	4 and 5	30299
7	exp autism spectrum disorder/di, dg, ep or Attention Deficit Disorder with Hyperactivity/di, dg, ep or "Attention Deficit and Disruptive Behavior Disorders"/di, dg, ep or Fetal Alcohol Spectrum Disorders/di, dg, ep	23876
8	6 or 7 [Developmental disorders]	42246
9	exp medical records/ or international classification of diseases/ or exp diagnostic techniques, neurological/ or exp clinical laboratory techniques/ or "diagnostic techniques and procedures"/	3106548
10	(((patient* or medical or health) adj3 record*) or ((diagnos* or defin* or classificat*) adj2 disease list) or ((self report* or standard*) adj2 measure*) or (icd adj (cod* or classification*)) or (international adj3 classification)).tw,kw,kf.	268618
11	medicaid/ or birth certificates/ or death certificates/ or hospital records/ or insurance claim reporting/ or exp insurance, health/ or databases, factual/ or information systems/ or databases as topic/ or database management systems/ or software/ or insurance claim review/ or patient discharge/ or exp registries/ or utilization review/	498782
12	(((administrat* or physician* or inpatient* or emergency* or hospital* or clinic or clinics or pharmac* or insurance) adj4 (admission* or data or dataset* or database* or data base? or data bank? or claim* or billing* or record* or utilizat* or utilisat*)) or ((claim* or discharg*) adj2 data*) or ((database? or databank? or data base? or data bank?) adj4 (factual or administrat* or claim? or register* or registr* or topic? or system?)) or (claim? adj2 (analy* or review? or physician? or pharmac* or drug?)) or (insurance adj2 (claim* or audit*)) or medicaid or (health* adj2 plan) or ((death* or birth*) adj1 (certificate* or record*))).tw,kw,kf. or (database? or databank? or data base? or data bank?).ti.	312830
13	or/9-12 [Diagnosis]	3955207
14	8 and 13	6161
15	exp Diagnostic Errors/ or Diagnosis, Differential/ or "Predictive Value of Tests"/ or "Sensitivity and Specificity"/ or ROC Curve/ or Area under Curve/ or Bayes Theorem/ or algorithms/ or validation studies as topic/	1317678
16	((diagnos* adj3 (schedul* or clinical* or technique* or procedur* or assess* or standard* or error* or false or incorrect* or wrong* or correct*)) or misdiagnos* or (clinical* adj (standard* or criteri* or measure* or classifi* or technique* or assess*)) or ((positive or negative) adj2 (predict* or false)) or sensitiv* or specif* or accura* or valid* or reliab* or agree* or concord* or misclass* or ((case or cases) adj2 ascertain*) or algorithm? or (bayes* adj (theorem or analysis or approach or forecast or method or prediction)) or (roc adj (curve or analysis)) or receiver operating characteristic).tw,kw,kf.	6045219
17	15 or 16	6721820
18	14 and 17	2712
19	limit 18 to (yr="1995-Current" and (english or french))	2429
20	(201908* or 201909* or 201910* or 201911* or 201912* or 202*).ez.	1099141
21	19 and 20	99

SECOND SEARCH UPDATE

Database(s): Ovid MEDLINE(R) ALL 1946 to March 30, 2021
Search Strategy:

#	Searches	Results
1	exp autism spectrum disorder/ or (autism or autistic or (asperger* adj (syndrome or disorder or disease)) or kanner* syndrome or childhood disintegrative disorder or (pervasive adj2 developmental disorder*) or heller* syndrome or disintegrative psychosis).tw,kf,kw.	57062
2	Attention Deficit Disorder with Hyperactivity/ or "Attention Deficit and Disruptive Behavior Disorders"/ or ((attention deficit adj4 disorder*) or (hyperkinetic adj2 (disorder or syndrome)) or minimal brain dysfunction or adhd or addh).tw,kf,kw.	44499
3	Fetal Alcohol Spectrum Disorders/ or (f?etal alcohol or (alcohol related adj3 (neurodevelopment or birth)) or "Neurobehavioral disorder associated with prenatal alcohol exposure" or "Growth Retardation, Facial Abnormalities, and Central Nervous System Dysfunction").tw,kw,kf.	6077
4	or/1-3	102829
5	diagnosis/ or incidence/ or prevalence/ or (diagnos* or incidence* or prevalence* or new case?).tw,kw,kf.	3904190
6	4 and 5	32263
7	exp autism spectrum disorder/di, dg, ep or Attention Deficit Disorder with Hyperactivity/di, dg, ep or "Attention Deficit and Disruptive Behavior Disorders"/di, dg, ep or Fetal Alcohol Spectrum Disorders/di, dg, ep	25017
8	6 or 7 [Developmental disorders]	44704
9	exp medical records/ or international classification of diseases/ or exp diagnostic techniques, neurological/ or exp clinical laboratory techniques/ or "diagnostic techniques and procedures"/	3171646
10	(((patient* or medical or health) adj3 record*) or ((diagnos* or defin* or classificat*) adj2 disease list) or ((self report* or standard*) adj2 measure*) or (icd adj (cod* or classification*)) or (international adj3 classification)).tw,kw,kf.	287276
11	medicaid/ or birth certificates/ or death certificates/ or hospital records/ or insurance claim reporting/ or exp insurance, health/ or databases, factual/ or information systems/ or databases as topic/ or database management systems/ or software/ or insurance claim review/ or patient discharge/ or exp registries/ or utilization review/	519079
12	(((administrat* or physician* or inpatient* or emergency* or hospital* or clinic or clinics or pharmac* or insurance) adj4 (admission* or data or dataset* or database* or data base? or data bank? or claim* or billing* or record* or utilizat* or utilisat*)) or ((claim* or discharg*) adj2 data*) or ((database? or databank? or data base? or data bank?) adj4 (factual or administrat* or claim? or register* or registr* or topic? or system?)) or (claim? adj2 (analy* or review? or physician? or pharmac* or drug?)) or (insurance adj2 (claim* or audit*)) or medicaid or (health* adj2 plan) or ((death* or birth*) adj1 (certificate* or record*))).tw,kw,kf. or (database? or databank? or data base? or data bank?).ti.	334466
13	or/9-12 [Diagnosis]	4068131
14	8 and 13	6575
15	exp Diagnostic Errors/ or Diagnosis, Differential/ or "Predictive Value of Tests"/ or "Sensitivity and Specificity"/ or ROC Curve/ or Area under Curve/ or Bayes Theorem/ or algorithms/ or validation studies as topic/	1352991
16	((diagnos* adj3 (schedul* or clinical* or technique* or procedur* or assess* or standard* or error* or false or incorrect* or wrong* or correct*)) or misdiagnos* or (clinical* adj (standard* or criteri* or measure* or classifi* or technique* or assess*)) or ((positive or negative) adj2 (predict* or false)) or sensitiv* or specif* or accura* or valid* or reliab* or agree* or concord* or misclass* or ((case or cases) adj2 ascertain*) or algorithm? or (bayes* adj (theorem or analysis or approach or forecast or method or prediction)) or (roc adj (curve or analysis)) or receiver operating characteristic).tw,kw,kf.	6345372
17	15 or 16	7034188
18	14 and 17	2895
19	limit 18 to (yr="1995-Current" and (english or french))	2611
20	(202007* or 202008* or 202009* or 20201* or 202*).ez.	1831745
21	19 and 20	182

Reporting quality assessment of included studies using modified STARD[a] checklist for validating health administrative data

Section, Topic and Item	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]
TITLE, KEYWORDS, ABSTRACT														
1. Identifies article as study of assessing diagnostic accuracy?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. Identifies article as study of administrative data?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes
INTRODUCTION														
3. States disease identification and validation as one of goals of study?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
METHODS														
Participants in validation cohort														
4. Describes validation cohort (cohort of patients to which reference standard was applied)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4a. Age?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4b. Disease?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4c. Severity?	No	No	No	No	No	No	No	No	No	No	No	No	No	No
4d. Location/jurisdiction?	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
5. Describes recruitment procedure of validation cohort?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5a. Inclusion criteria?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5b. Exclusion criteria?	Yes	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes
6. Describes patient sampling (random, consecutive, all, etc.)?	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
7. Describes data collection?	n/a	Yes	Yes	Yes	Yes	Yes	n/a	Yes	Yes	Yes	Yes	Yes	Yes	Yes
7a. Who identified patients and ensured selection adhered to patient recruitment criteria?	n/a	Yes	Yes	Yes	Yes	Yes	n/a	Yes	Yes	Yes	Yes	Yes	Yes	Yes
7b. Who collected data?	n/a	Yes	Yes	Yes	Yes	Yes	n/a	Yes	Yes	Yes	Yes	Yes	Yes	Yes
7c. *A priori* data collection form?	n/a	No	Yes	Yes	Yes	Yes	n/a	No	Uncertain	Yes	Yes	No	Yes	No
7d. How was disease classified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
8. Was there a split sample (i.e. re-validation using a separate cohort)?	No	Yes	No[b]	No	No	No	No	No	No	No	No	No	No	No

Section, Topic and Item	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]
Test methods														
9. Describe number, training and expertise of persons reading reference standard?	n/a	Yes	Yes	Yes	Yes	No	n/a	No	Yes	Yes	Yes	Yes	Yes	No
10. If >1 person reading reference standard, measure of consistency is reported (e.g. kappa)?	n/a	No	No	Yes	Yes	n/a	n/a	No	Yes	n/a	No	Yes	Yes	n/a
11. Were the readers of the reference (validation) test blinded to the results of the classification by administrative data for that patient? (e.g. Was the reviewer of the charts blinded to how that chart was billed?)	n/a	Yes	Yes	Uncertain	No	Yes	n/a	No	No	No	No	No	No	Yes
Statistical methods														
12. Describe methods of calculating/ comparing diagnostic accuracy?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
RESULTS														
Participants														
13. Report when study done, start/end dates of enrolment?	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No
14. Describe number of people who satisfied inclusion/exclusion criteria?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
15. Study flow diagram?	No	Yes	Yes	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes	No
Test results														
16. Reports distribution of disease severity?	No	No	No	No	No	No	No	No	No	No	No	No	No	No
17. Report cross-tabulation of index tests by results of reference standard?	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes

Continued on the following page

Reporting quality assessment of included studies using modified STARD[a] checklist for validating health administrative data

Section, Topic and Item	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]
Estimates														
18. Reports at least 4 estimates of diagnostic accuracy? (Estimates reported in included studies)	Yes	Yes	Yes	No	No	No	No	No	No	No	No	Yes	No	No
18a. Sensitivity	Yes	Yes	Yes	No	No	Yes	Yes	No	No	No	No	Yes	No	No
18b. Specificity	Yes	Yes	Yes	No	No	No	Yes	No	No	No	No	Yes	No	No
18c. PPV	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
18d. NPV	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	Yes	No	Yes
18e. Likelihood ratios	No	No	No	No	No	No	No	No	No	No	No	No	No	No
18f. Kappa	Yes	No	No	No	No	No	No	No	No	No	No	Yes	No	No
18g. Area under the ROC curve / c-statistic	Yes	No	No	No	No	No	Yes	No	No	No	No	No	No	No
18h. Accuracy/agreement	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No
19. Was the accuracy reported for any subgroups (e.g. age, geography, different sex etc.)?	No	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No
20. If PPV/NPV reported, does ratio of cases/controls of validation cohort approximate prevalence of condition in the population?	No	Yes	Yes	No	No	No	n/a	No	No	No	No	No	No	No
21. Reports 95% CIs for each diagnostic accuracy measure?	Yes	Yes	Yes	No	No	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes
DISCUSSION														
22. Discusses the applicability of the findings?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Abbreviations: CI, confidence interval; n/a, not applicable; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic.

[a] Standards for the Reporting of Diagnostic Accuracy Studies (STARD): method of assessing reporting quality of validation studies using administrative data.

[b] This study re-validated one algorithm using a separate cohort; however, the algorithm did not include health administrative data and therefore was out of scope for this review.

Study Domains	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. PATIENT SELECTION														
A. Risk of Bias														
Q1	Yes	Yes	Yes	No	Unclear	No[b] Yes[c]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Q2	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Q3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	No
Risk	**HIGH**	**LOW**	**LOW**	**HIGH**	**UNCLEAR**	**HIGH[b] LOW[c]**	**LOW**	**LOW**	**LOW**	**HIGH**	**HIGH**	**HIGH**	**LOW**	**HIGH**
B. Applicability Concerns														
Concern	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**
2. INDEX TEST														
A. Risk of Bias														
Q1	Unclear	Unclear	Unclear	Yes	Yes	No	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Risk	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**
B. Applicability Concerns														
Concern	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**
3. REFERENCE STANDARD														
A. Risk of Bias														
Q1	Yes	Yes	Yes	Yes	Yes	Unclear[b] No[c]	Yes	Unclear	Yes	Yes	Yes	No	Yes	Unclear
Q2	Yes	Yes	Yes	Unclear	No	Yes[b] No[c]	Yes	No	No	No	No	No	No	Yes
Risk	**LOW**	**LOW**	**LOW**	**UNCLEAR**	**LOW**	**UNCLEAR[b] HIGH[c]**	**LOW**	**UNCLEAR**	**LOW**	**LOW**	**LOW**	**HIGH**	**LOW**	**UNCLEAR**
B. Applicability Concerns														
Concern	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**	**LOW**
4. FLOW AND TIMING														
Risk of Bias														
Q1	Yes	Unclear	Unclear	Yes	Unclear	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Unclear
Q2	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Q3	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No	Yes	Yes	No	Yes
Risk	**LOW**	**LOW**	**LOW**	**LOW**	**HIGH**	**HIGH**	**HIGH**	**LOW**	**HIGH**	**HIGH**	**LOW**	**HIGH**	**HIGH**	**LOW**

[a] Revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool: method of assessing the risk of bias and applicability of diagnostic studies through four key study domains, with each domain rated as low, high or unclear risk of bias and low, high or unclear applicability to the research question.

[b] The risk of bias for the portion of the study evaluating sensitivity using the study's "sensitivity cohort".

[c] The risk of bias for the portion of the study evaluating positive predictive value using children with an administrative diagnosis of autism spectrum disorder.

Risk of bias and applicability concerns: Signalling questions and scoring guidelines

Study Domains	
1. PATIENT SELECTION	
A. Risk of Bias	
Q1	**Was a consecutive or random sample of patients enrolled?** Yes/No/Unclear • Select 'Yes' if consecutive or random sampling was used to select patients for the validation cohort. • Select 'No' if non-consecutive or convenience sampling was used. • Select 'Unclear' if insufficient information is reported.
Q2	**Was a case-control design avoided?** Yes/No/Unclear • Select 'Yes' if a case-control design was avoided. • Select 'No' if patients were selected based on known disease (i.e., confirmed as opposed to suspected cases) and non-disease status. • Select 'Unclear' if insufficient information is reported.
Q3	**Did the study avoid inappropriate exclusions?** Yes/No/Unclear • Select 'Yes' if the study avoided inappropriate exclusions. • Select 'No' if the study excluded patients inappropriately, such as excluding difficult to diagnose patients or suspected but unconfirmed diagnoses. • Select 'Unclear' if insufficient information is reported.
Risk	**Could the selection of patients have introduced bias?** LOW/HIGH/UNCLEARᵃ
B. Applicability Concerns	
Concern	**Is there concern that the included patients do not match the review question?** LOW/HIGH/UNCLEAR
2. INDEX TEST	
A. Risk of Bias	
Q1	**Were the administrative database algorithm(s) results interpreted without knowledge of the results of the reference standard?** Yes/No/Unclear • Select 'Yes' if the algorithm(s) results were interpreted without knowledge of the reference standard diagnosis. • Select 'No' if it was reported that the algorithm(s) results were interpreted with knowledge of the results of the reference standard diagnosis. • Select 'Unclear' if insufficient information is reported.
Risk	**Could the conduct or interpretation of the algorithm(s) have introduced bias?** LOW/HIGH/UNCLEARᵃ
B. Applicability Concerns	
Concern	**Is there concern that the algorithm(s), its/their conduct or interpretation differ from the review question?** LOW/HIGH/UNCLEAR
3. REFERENCE STANDARD	
A. Risk of Bias	
Q1	**Is the reference standard likely to correctly classify the target condition?** Yes/No/Unclear • Select 'Yes' if established clinical classification criteria, clinical case definitions derived from medical records or a medical record diagnosis was used; if experienced or trained personnel carried out the record review/ abstractions (where applicable); and if agreement was calculated to be high when more than one person reviewed/abstracted data. • Select 'No' if the reference standard was patient self-report; the personnel reviewing/abstracting information from the reference standard had insufficient experience or training (where applicable); or in cases where more than one person reviewed or abstracted data, agreement between personnel was low. • Select 'Unclear' if insufficient information is reported (e.g., no information was reported on interrater agreement when more than one person reviewed).

Study Domains	
Q2	**Were the reference standard results interpreted without knowledge of the results of the algorithm(s)?** Yes/No/Unclear • Select 'Yes' if the reference standard results were interpreted without knowledge of the algorithm(s) results. • Select 'No' if the reference standard was applied with knowledge of the algorithm(s) results, including when only patients flagged by the algorithm(s) received the reference standard. • Select 'Unclear' if insufficient information is reported.
Risk	**Could the reference standard, its conduct or its interpretation have introduced bias?** LOW/HIGH/UNCLEAR^a
B. Applicability Concerns	
Concern	**Is there concern that the target condition as defined by the reference standard does not match the review question?** LOW/HIGH/UNCLEAR
4. FLOW AND TIMING	
Risk of Bias	
Q1	**Was there an appropriate interval between ascertaining cases from the algorithm(s) and the reference standard?** Yes/No/Unclear • Select 'Yes' if there was an appropriate time interval between the algorithm(s) and reference standard. • Select 'No' if the time period between the reference standard diagnosis and algorithm(s) diagnosis was not appropriate. • Select 'Unclear' if insufficient information is reported.
Q2	**Did patients receive the same reference standard?** Yes/No/Unclear • Select 'Yes' if patients received the same reference standard. • Select 'No' if different reference standards were used. • Select 'Unclear' if insufficient information is reported.
Q3	**Were all patients included in the analysis?** Yes/No/Unclear • Select 'Yes' if the number of patients enrolled (i.e., after exclusions) is the same as the number of patients included in the 2x2 table of results. • Select 'No' if the number of patients enrolled differs from the number of patients included in the 2x2 table of results. • Select 'Unclear' if insufficient information is reported (e.g., no information on how the final validation study population was achieved).
Risk	**Could the patient flow have introduced bias?** LOW/HIGH/UNCLEAR^a

^a **Scoring guidelines:**

- If answers to all signalling questions within a domain were "yes" then risk of bias was judged as "LOW".
- If answers to all signalling questions within a domain were "no" then risk of bias was judged as "HIGH".
- If answers to all signalling questions within a domain were "unclear" then risk of bias was judged as "UNCLEAR".
- If any one signalling question was "no" this flagged the *potential* for bias and the review authors decided on what basis a judgment of high risk of bias might be made under such circumstances.
  ‣ The signalling question for "Index Test" (Q1) was considered a less important source of bias for this review. The second signalling question for "Reference Standard" (Q2) was also considered a less important source of bias but a judgment was made on a study-by-study basis. For all other signalling questions, one "no" response was sufficient for a judgment of high risk of bias.
- If any one signalling question was "unclear" this flagged the *potential* for bias and the review authors decided on what basis a judgment of unclear risk of bias might be made under such circumstances.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. Washington (DS): American Psychiatric Publishing Inc; 2013. 991 p.

2. Langlois K, Samokhvalov A, Rehm J, Spence S, Conner Gorber S. Health state descriptions for Canadians: mental illnesses. Ottawa (ON): Statistics Canada; 2011 [Statistics Canada, Catalogue No.: 82-619-MIE2005002].

3. Chudley AE, Conry J, Cook JL, Loock C, Rosales T, LeBlanc N. Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis. CMAJ 2005;172(5 Suppl):S1-S21. https://doi.org/10.1503/cmaj.1040302

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;10:283-92. 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):352-8. 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):323-37. 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, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. J Clin Epidemiol. 2011;64(8):821-9 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. 217 p.

14. Sørensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. Int J Epidemiol. 1996;25(2):435-42. 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):1446-52. 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. 238 p.

19. Shortliffe E, Cimino J, editors. Biomedical informatics: computer applications in health care and biomedicine. 3rd ed. New York (NY): Springer; 2006. 2100 p.

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):529-36. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

21. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York (UK): University of York; 2009. 294 p.

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):456-63. 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):1037-45. 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):400-10. 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):321-30. 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):1989-96. 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 database-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. Validity of autism diagnoses using administrative health data. Chronic Dis Can. 2009;29(3):102-7.

29. Hagberg KW, Jick SS. Validation of autism spectrum disorder diagnoses recorded in the Clinical Practice Research Datalink, 1990-2014. Clin Epidemiol. 2017;9:475-82. 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):139-48. 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):416-25. 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):1647-55. 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):191-7. 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):169-76. 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):1490-503. 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