Health Promotion and Chronic Disease Prevention in Canada

Research, Policy and Practice

Volume 42 · Number 9 · September 2022

Inside this issue

Indexed in Index Medicus/MEDLINE, DOAJ, SciSearch® and Journal Citation Reports/Science Edition

Public Health Agency of Canada | Agence de la santé publique du Canada

Canada

To promote and protect the health of Canadians through leadership, partnership, innovation and action in public health.
— Public Health Agency of Canada

Published by authority of the Minister of Health.

ISSN 2368-738X
Pub. 210491
PHAC.HPCDP.journal-revue.PSPMC.ASPC@canada.ca

Également disponible en français sous le titre : *Promotion de la santé et prévention des maladies chroniques au Canada : Recherche, politiques et pratiques*

Submission guidelines and information on article types are available at:
https://www.canada.ca/en/public-health/services/reports-publications/health-promotion-chronic-disease-prevention-canada-research-policy-practice/information-authors.html

https://doi.org/10.24095/hpcdp.42.9.01

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.

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]

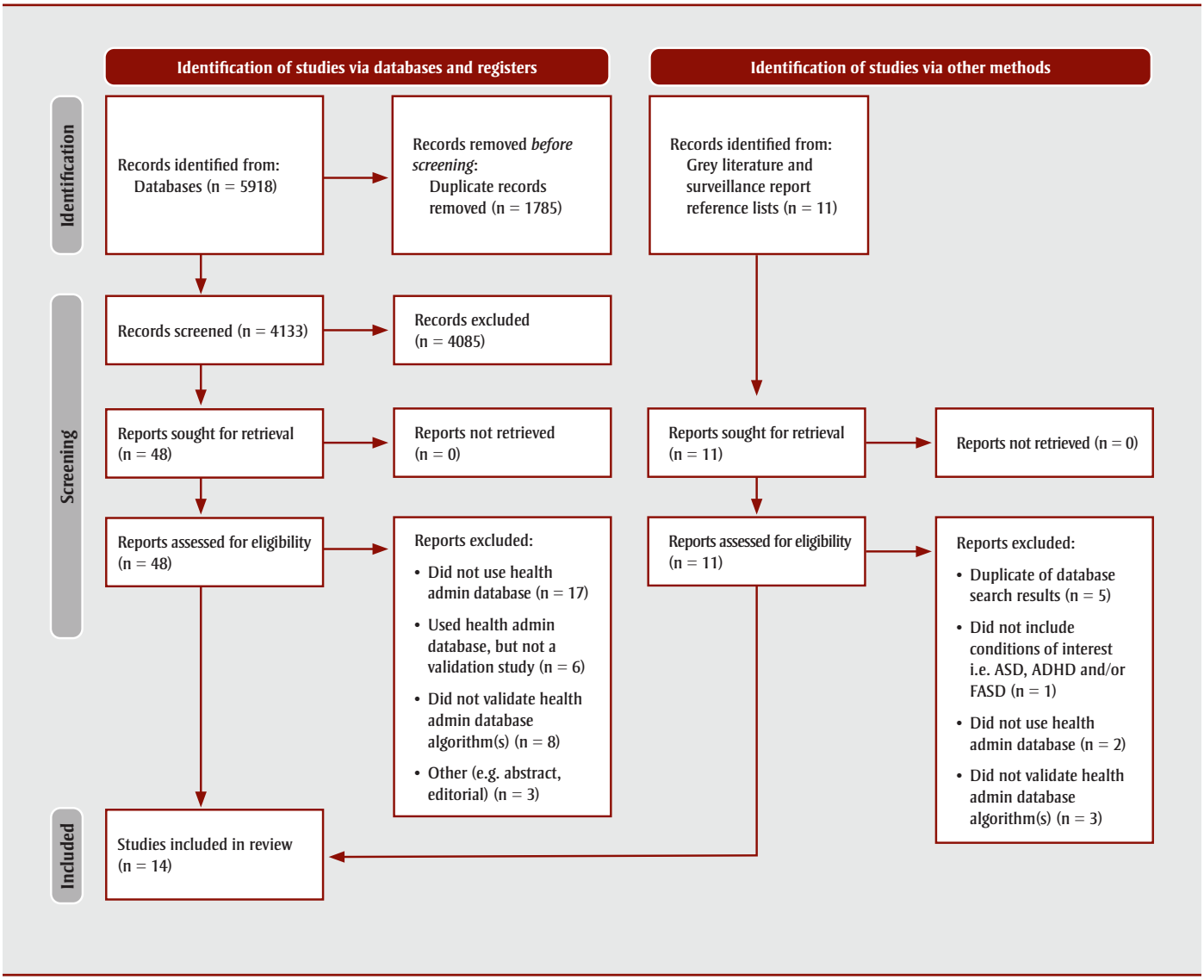
The PPV was the most commonly reported measure and was reported in 9 of the 10 studies.[22-27,29-31] Only three studies reported at least four measures of diagnostic accuracy.[22-24]

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

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)
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

TABLE 2
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]
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)

Continued on the following page

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

FIGURE 2
Risk of bias and applicability concerns of included studies by QUADAS-2[a] domain

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%)

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]
		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

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]
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]

Acknowledgements

The authors wish to acknowledge Katherine Merucci, a reference librarian from the Health Library within the Corporate Services Branch of Health Canada and Public Health Agency of Canada, who developed and ran the database and grey literature search strategies.

No funding, including grants or other research support, was obtained for this study.

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.

APPENDIX A
Electronic search strategy

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

APPENDIX A (continued)
Electronic search strategy

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

APPENDIX A (continued)
Electronic search strategy

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

APPENDIX B
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

Continued on the following page

APPENDIX B (continued)
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]
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

APPENDIX B (continued)
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.

APPENDIX C
Risk of bias and applicability assessments using QUADAS-2[a] tool

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.

APPENDIX C (continued)
Risk of bias and applicability assessments using QUADAS-2[a] tool

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/UNCLEAR[a]
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/UNCLEAR[a]
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).

Continued on the following page

APPENDIX C (continued)
Risk of bias and applicability assessments using QUADAS-2[a] tool

Study Domains	
	Were the reference standard results interpreted without knowledge of the results of the algorithm(s)? Yes/No/Unclear
Q2	• 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	
	Was there an appropriate interval between ascertaining cases from the algorithm(s) and the reference standard? Yes/No/Unclear
Q1	• 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.
	Did patients receive the same reference standard? Yes/No/Unclear
Q2	• Select 'Yes' if patients received the same reference standard. • Select 'No' if different reference standards were used. • Select 'Unclear' if insufficient information is reported.
	Were all patients included in the analysis? Yes/No/Unclear
Q3	• 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 data-based case definition for identifying children and youth with autism spectrum disorder for surveillance purposes. Can J Public Health. 2017;108(5-6): e530-e538. https://doi.org/10.17269/cjph.108.5963

28. Dodds L, Spencer A, Shea S, et al. 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

https://doi.org/10.24095/hpcdp.42.9.02

Evidence synthesis

Sleep duration and eating behaviours among adolescents: a scoping review

Natalie Doan, MSc (1); Amanda Parker, MSc (1); Katherine Rosati, BSc (2); Ella van Beers, BScH (3); Mark A. Ferro, PhD (1)

This article has been peer reviewed.

Abstract

Introduction: In the past decade, investigations of the relationship between sleep duration and eating behaviours have been emerging; however, a formal synthesis of the literature focussed on adolescent populations has not yet been conducted. We conducted a scoping review of the literature examining the relationship between sleep duration and eating behaviours in adolescents. Gaps in the research and directions for future research were identified based on the findings.

Methods: A systematic search was employed on four research databases: PubMed, PsycInfo, CINAHL and Scopus; relevant grey literature was also reviewed. Studies that reported on the relationship between sleep duration and eating behaviours among high school–aged adolescents were included in the review. Data were extracted, charted and synthesized into a narrative. Consistent with the purpose of a scoping review, the methodological quality of the studies was not appraised. Stakeholders were consulted to validate the findings and provide insight into the interpretation and identification of pressing gaps in the research that remain to be addressed.

Results: In total, 61 studies published between 2006 and 2021 met the criteria for review. Existing research focussed heavily on examining sleep duration in relation to intake of food from certain food groups, beverages and processed foods, and relied on a population study design, cross-sectional analyses and self-report measures.

Conclusion: Future research is needed to understand the link between sleep duration and eating-related cognition, eating contexts and disordered eating behaviours in order to better understand how ensuring sufficient sleep among adolescents can be leveraged to support healthier eating practices and reduce diet-related risks.

Keywords: sleep, dietary patterns, eating habits, youth, adolescents

Highlights

- Unique to this study, we reviewed the breadth of the literature related to sleep duration, dietary intake and eating habits among adolescents.
- We found a large emphasis on the dietary intake of healthful foods, beverages and processed foods, and limited focus on the contextual factors that shape eating, eating-related cognitions and disordered eating symptoms.
- Stakeholders validated the findings, provided insight into the interpretation of the findings and highlighted areas for future research.
- Additional sleep research exploring the cognitive and contextual factors surrounding eating is needed (e.g. eating with others, eating when not hungry, binge eating).

Introduction

Evidence suggests that poor dietary patterns characterized by excessive intake of sugar, saturated fat and salt, as well as low intake of vegetables and whole grains, are associated with the development of noncommunicable diseases (e.g. diabetes).[1,2] The role of psychosocial eating habits, such as eating with other people and mindful eating, is also being increasingly recognized for its role in supporting healthy eating.[3,4] Hence, supporting healthy food consumption and eating habits is critical to prevent and manage diet-related diseases. Given the role of sleep in regulating hormones that affect appetite (e.g. insulin, leptin, ghrelin),[5-8] fostering adequate sleep among adolescents could support the development of a range of healthy behaviours during adolescence.[9-11]

Considering the importance of the adolescent period in the development of life-long behaviours, promoting healthy eating behaviours among adolescents is critical.[12] Adolescence is especially important because it is characterized by many developmental and behavioural changes, including a decline in healthy eating habits.[13] Additionally, changes to the circadian rhythm that occur during this developmental period result in a natural shift towards later sleep onset among adolescents.[14] This shift in the circadian rhythm can contribute to insufficient sleep, which is further compromised by changes such as early school start times, increased academic

Author references:

1. School of Public Health Sciences, University of Waterloo, Waterloo, Ontario, Canada
2. Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada
3. Department of Psychology, University of Toronto, Toronto, Ontario, Canada

Correspondence: Natalie Doan; 200 University Avenue West, Waterloo, ON N2L 3G1; Tel: 519-888-4567; Email: natalie.doan@uwaterloo.ca

demands and extracurricular activities.[15-17] Therefore, examining the relationship between these two modifiable lifestyle factors (sleep duration and eating behaviours) among adolescents is necessary to better understand approaches to facilitating health and the development of healthy life-long behaviours.

Sleep is an essential component of healthy development during adolescence. For optimal health and well-being, it is recommended that adolescents aged 15 to 17 years get 8 to 10 hours of sleep per night.[18] In the past decade, adequate sleep as a lifestyle factor has garnered increasing attention in the literature. Short sleep duration over a prolonged period of time is associated with a range of adverse physical and emotional health outcomes (e.g. mood dysregulation, accidental injuries).[15,19] Insufficient sleep, in particular, has been associated with poor dietary intake and the development of diet-related diseases;[13,20] potentially in part due to alterations in metabolic hormone regulation, as well as extended waking hours.[5,11,21] Despite these potential links, little is known about the generalizability of the relationship between eating behaviours and sleep duration in adolescent populations. Thus, understanding this relationship is crucial for clinicians and researchers to better understand the complex relationships among sleep duration and diet-related diseases.[22]

The objectives of this scoping review were to systematically review the literature that examines sleep duration in relation to eating behaviours among adolescents, and to identify gaps and provide direction for future research in the field of adolescent health promotion.

Methods

This review follows the six-staged framework for scoping studies described by Arksey and O'Malley[23] and recommendations outlined by Levac and colleagues[24] to enhance the scoping review methodologies. The six stages of this framework are: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; (5) collating, summarizing and reporting the results; and (6) consultation exercise.[23] The protocol for this study is available in further detail elsewhere.[25] This review is reported in accordance with the PRISMA Extension for Scoping Review (PRISMA-ScR) guidelines.[26,27]

Identifying the research questions

The primary research question guiding this scoping review was: what is the nature of the research on the relationship between sleep duration and eating behaviours in adolescents? Grounded in the objectives of performing a scoping review—to map the key concepts and evidence available on a research topic—the secondary research questions were: (1) which research designs have been employed?; (2) which adolescent populations have been studied?; (3) which outcome variables have been assessed?; and (4) what questions remain to be addressed? Methodological quality of the studies was not assessed, given the primary purpose of conducting a scoping review.[23]

Identifying relevant studies

To identify relevant studies, systematic searches were conducted on PubMed, CINAHL, PsycInfo, and Scopus. The most recent search strategy (Table 1) was executed on 17 November 2021. Grey literature studies were also reviewed using the template provided by Godin et al.[28] Using this guideline, a targeted search of relevant health organization websites and public health databases was conducted. The grey literature search was conducted on 20 March 2020.

Selecting the studies

The studies were screened to ensure that findings reported on the association between sleep duration and eating behaviours among adolescents (approximately 13–19 years of age). No restrictions were placed on research approaches, study design or study type. Studies that solely examined infants, toddlers, preschoolers, school-aged children, adults or older adults were excluded. Only studies reported in English and in the form of a publication, thesis, dissertation, technical report or conference proceedings were included in the final review.

Studies were screened using a two-level screening process to determine eligibility. Studies were first reviewed based on the title and abstract, and then selected for inclusion in the final review based on reading the article in full. In the first stage, the title and abstract of each study were independently screened by one reviewer to determine potential eligible studies. In the second stage, the full text of each study was screened by two reviewers, after which the reviewers met to discuss the cases for which the decision was not unanimous, in order to reach consensus.

Charting the data

The data extraction stage followed the two-step recommendations by Daudt et al.[29] To ensure the validity of the data extracted, the review team met to discuss the data extraction protocol. Following that, each member of the review team independently extracted data from articles that were purposely chosen to represent a range of themes and study designs. After independently extracting data from the same set of studies, the review team reconvened to discuss discrepancies before independently extracting data from the remaining studies. Key characteristics were extracted and recorded using a spreadsheet, including publication, study and population characteristics. Data pertaining to the research focus on sleep duration and eating behaviours were extracted. The first author reviewed the data extracted from all studies for accuracy.

Collating, analyzing and synthesizing the data

The charted data were collated, analyzed and synthesized to summarize the current body of literature on sleep duration and eating behaviours in adolescent populations. This summary is presented in the form of aggregate numeric values and narrative descriptions in the results section. The findings are grouped under two primary domains: food consumption and eating habits.

Consulting stakeholders

Based on the grey literature review, three stakeholders were identified, and two were contacted via email. These two stakeholders were specifically consulted based on their content expertise (e.g. adolescent health, eating- and weight-related behaviours) and profession. One was a researcher and clinician, and the other was a clinician and community worker. Both stakeholders agreed to participate in a consultation for this scoping review. The first stakeholder was a youth health researcher, with frontline clinical experiences, specializing in population-level primary prevention and health promotion research. The second stakeholder was an education and outreach coordinator and

TABLE 1
Keywords and search terms employed in the systematic search

Study population	Sleep duration	Eating behaviour
Youth	Sleep duration	Food
Adolescent	Sleep quantity	Diet
Adolescence	Sleep deprivation	Dietary
Teenager	Insufficient sleep	Dieting
Teen	Sleep restriction	Nutrition
	Deprived sleep	Weight control
	Restricted sleep	Eating behaviour
	Disrupted sleep	Eating
	Sleep disruption	Feeding behaviour
	Excessive sleep	Feeding
	Over sleep	Calorie
	Sleep disorders	Vegetables
	Insomnia	Fruit
	Hypersomnia	Soda
		Carbohydrates
		Fats
		Milk
		Proteins
		Snacks
		Sugar
		Obesity
		Binge
		Bingeing
		Anorexia
		Bulimia

psychotherapist, for a national organization that delivers community education and school-based prevention programming. During the consultations, the first author shared preliminary findings and validated the interpretations with the stakeholders. Stakeholders were solicited for their perspectives on important directions for future research in the field and provided relevant articles for the team to review. The perspectives gathered through the consultation exercise guided the reporting and interpretation of the results. The areas identified as priorities for future research were used to frame the discussion of the results.

Results

Study selection

The systematic and grey literature searches yielded 2185 and 106 citations, respectively. After removing duplicates, the remaining 2291 citations were screened. A total of 61 articles from the systematic and grey literature searches met eligibility criteria and were included in the final synthesis (Figure 1).

Publication, study and population characteristics

Tables 2 and 3 present the study characteristics of the peer-reviewed and grey literature studies included in the final review, respectively. All studies were published between 2006 and 2021. Most of the studies were conducted in North America (36.1%), Europe (23.0%) and East Asia (13.1%). The sample size of the included studies ranged from 21 to 1 777 091, with a median sample size of 1522.

Table 4 presents the study and population characteristics of the 61 included studies. Most studies published on this topic used a cross-sectional design (86.9%), were observational in nature (93.4%) and took place in a school setting (57.4%). Many studies exclusively examined adolescents within the high school age range (41.0%); however, some also included younger (50.8%) or older (3.3%) adolescents in their sample. With the exception of four (and three that did not specify), most studies examined both males and females (88.5%).

The published research in this domain predominantly used self-report measures of sleep duration; 72.1% of the studies included in our review used such self-report measures (e.g. questionnaire, sleep diary, interview recall, guardian report). Objective measures of sleep duration were used in 16.4% of studies (e.g. actigraphy, accelerometer, polysomnography) and a combination of self-report and objective measures of sleep duration were used in 9.8% of the studies (data not shown).

Most studies examined multiple aspects of eating behaviours using self-report measures. In 72.1% of studies, self-report questionnaires were used and in 14.8% of studies, interview methods were used to obtain a measure of dietary intake. In 6.6% of studies, objective measures of dietary intake, such as analysis of meal orders and absolute caloric intake, were used. Two studies assessed eating behaviours using experimental tasks (3.3%). One study used food records (1.6%), and another used a combination of 24-hour recall and food records (1.6%).

Research focus on sleep duration and eating behaviour studies

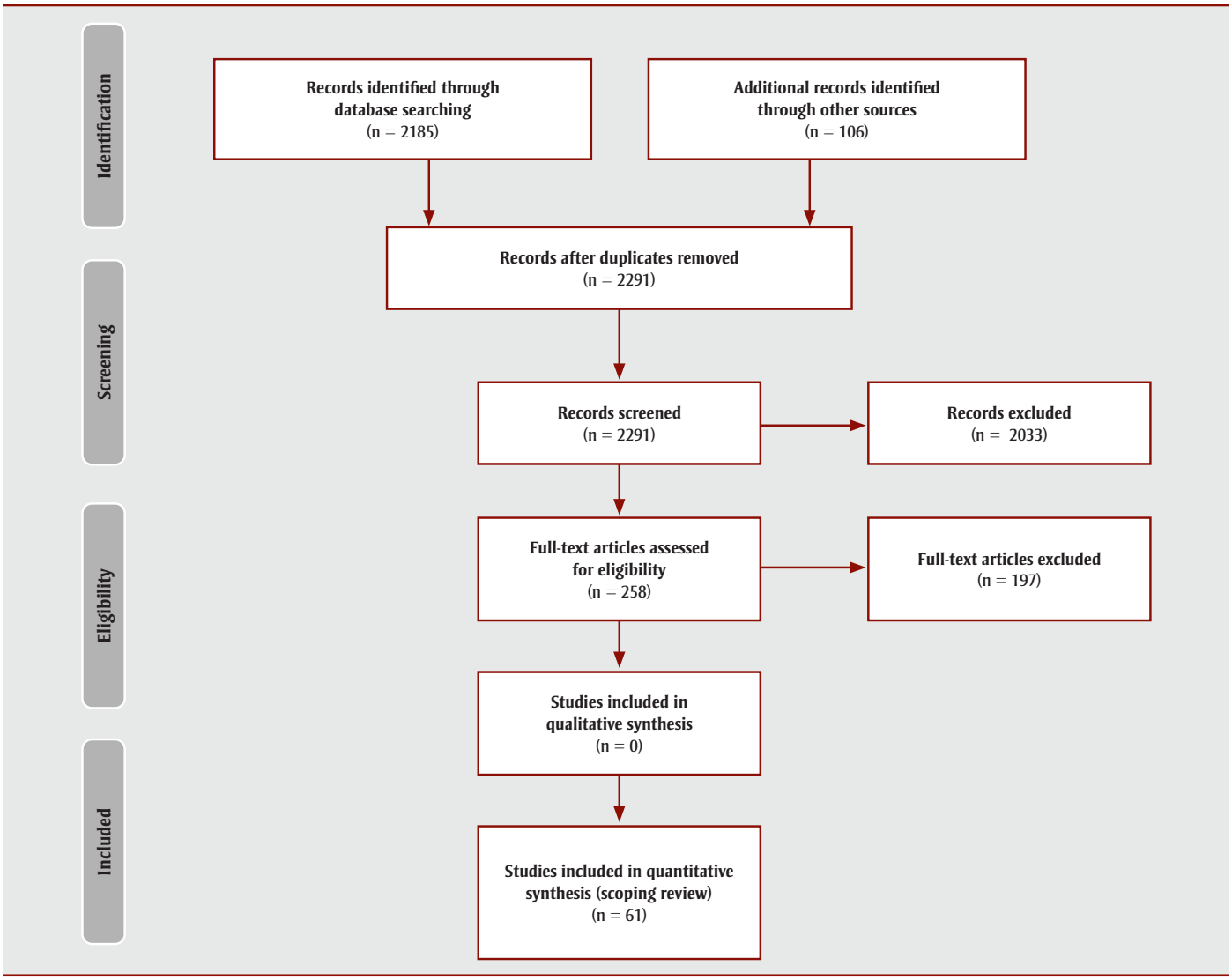
Table 5 presents the frequency with which the research foci were reported in the literature.

Food consumption

Food group intake

Indicators of food groups were represented in the largest number of studies. The food group most commonly assessed was fruit and vegetables. Among studies that examined fruit and vegetable intake, findings were mixed. Ten studies found that vegetable consumption was positively associated with sufficient sleep.[30-39] One study found that higher intake of fruit and vegetables was associated with shorter self-reported sleep duration.[40] One study found that longer sleep duration was associated with higher fruit and vegetable consumption in males, but not females.[41] Two studies did not reveal any significant associations between fruit and vegetable

FIGURE 1
PRISMA flow chart[a]



[a] Based on the guidelines summarized by Moher et al.[27]

consumption and sleep duration.[42,43] Eight studies reported on intake of milk and dairy products, meat and alternatives and grain products, and the findings were mixed.[30,31,33,35,42,44,45,46]

Beverage intake

Beverage intake was the second most frequently examined variable. A number of studies observed that short sleep was associated with greater intake of sugar-sweetened beverages[31,35,36,42,47] and soft drinks.[44,48,49] One study observed that short sleep was associated with lower odds of intake of soft drinks without sugars.[35]

Processed food intake

Studies frequently reported on intake of processed foods, including fast food, sweets and salty snacks. Among these studies, it was often reported that short sleep was associated with higher consumption of fast food,[34,42] sweets[32,42,44,50] and salty snacks,[32,35,51] with the exception of one study that did not find a significant association between sleep duration and fast food consumption.[38]

Energy intake

Energy intake, or caloric intake, was a common indicator examined among the studies. Studies reported mixed findings regarding the direction and significance of the relationship between sleep duration and energy intake. Some studies reported findings that suggest that short sleep duration was associated with higher energy intake,[52,53] while one study reported short sleep duration was associated with a small negative energy balance,[54] and two reported insignificant findings.[55,56]

Caffeinated beverage intake

Nine studies examined the associations between sleep duration and caffeinated beverages.[35,36,42,47,49,51,57-59] Of these, four reported significant findings that consumption of caffeinated beverages was associated with shorter sleep duration.[47,49,57,58] One study found that short sleep was associated with a decreased intake of coffee.[35]

Macronutrient intake

Intake of macronutrients was reported in eight studies.[56,60-66] Using 24-hour-food-recall questionnaires and wrist-actigraphy measures, one study found that those who

TABLE 2
Publication characteristics of the studies included from the systematic search

No.	Title	Authors	Year	Country
1	The influence of adherence to the Mediterranean diet on academic performance is mediated by sleep quality in adolescents	Adelantado-Renau et al.[68]	2019	Spain
2	Subjective sleep duration and quality influence diet composition and circulating adipocytokines and ghrelin levels in teen-age girls	Al-Disi et al.[61]	2010	Saudi Arabia
3	Relative contribution of obesity, sedentary behaviors, and dietary habits to sleep duration among Kuwaiti adolescents	Al-Haifi et al.[42]	2016	Kuwait
4	Lifestyle correlates of self-reported sleep duration among Saudi adolescents: a multicentre school-based cross-sectional study	Al-Hazzaa et al.[43]	2014	Saudi Arabia
5	The association between obstructive sleep apnea and dietary choices among obese individuals during middle to late childhood	Beebe et al.[62]	2011	United States
6	Psychiatric morbidity and dietary habits during COVID-19 pandemic: a cross-sectional study among Egyptian youth (14–24 years)	Alamrawy et al.[59]	2021	Egypt
7	Association of overweight, obesity and insufficient sleep duration and related lifestyle factors among school children and adolescents	Almulla and Zoubeidi[49]	2021	United Arab Emirates
8	Association between food patterns and difficulties in falling asleep among adolescents in Norway — a descriptive Young-Hunt3 study	André et al.[46]	2021	Norway
9	Association between self-reported sleep duration and dietary quality in European adolescents	Bel et al.[67]	2013	Germany, Belgium, Italy, France, Spain, Austria, Greece, Sweden, United Kingdom
10	Short sleep duration is associated with specific food intake increase among school-aged children in China: a national cross-sectional study	Cao et al.[31]	2019	China
11	How do energy balance-related behaviors cluster in adolescents?	Collese et al.[90]	2018	Maringa (Brazil), Athens (Greece), Dortmund (Germany), Ghent (Belgium), Heraklion (Greece), Lille (France), Pecs (Hungary), Rome (Italy), Stockholm (Sweden), Vienna (Austria) and Zaragoza (Spain)
12	The impact of short sleep on food reward processes in adolescents	Duraccio et al.[91]	2019	United States
13	Effects of sleep restriction on food-related inhibitory control and reward in adolescents	Duraccio et al.[76]	2019	United States
14	Sleep and pre-bedtime activities in New Zealand adolescents: differences by ethnicity	Galland et al. [58]	2020	New Zealand
15	Short sleep duration is associated with increased obesity markers in European adolescents: effect of physical activity and dietary habits. The HELENA study	Garaulet et al.[33]	2011	Austria, Belgium, France, Germany, Greece, Hungary, Italy, Spain, Sweden
16	A chrononutrition perspective of diet quality and eating behaviors of Brazilian adolescents in associated with sleep duration	Garcez et al. [72]	2021	Brazil
17	Sleep duration or bedtime? Exploring the association between sleep timing behaviour, diet and BMI in children and adolescents	Golley et al.[52]	2013	Australia
18	Family dinner frequency is inversely related to mental disorders and obesity in adolescents: the CASPIAN-III study	Haghighatdoost et al.[73]	2017	Iran
19	Sleep patterns and quality are associated with severity of obesity and weight-related behaviors in adolescents with overweight and obesity	Hayes et al.[55]	2018	United States
20	Racial/ethnic disparity in habitual sleep is modified by caloric intake in adolescents	He et al.[64]	2020	United States
21	Habitual sleep variability, not sleep duration, is associated with caloric intake in adolescents	He et al.[56]	2015	United States
22	Behaviors associated to sleep among high school students: cross-sectional and prospective analysis	Hoefelmann et al.[74]	2014	Brazil
23	Association between unhealthy behavior and sleep quality and duration in adolescents	Hoefelmann et al.[92]	2015	Brazil
24	Dietary intake and eating-related cognitions related to sleep among adolescents who are overweight or obese	Ievers-Landis et al.[63]	2016	United States

Continued on the following page

TABLE 2 (continued)
Publication characteristics of the studies included from the systematic search

No.	Title	Authors	Year	Country
25	Dietary patterns in relation to prospective sleep duration and timing among Mexico City adolescents	Jansen et al.[45]	2020	Mexico
26	Relationships of beverage consumption and actigraphy-assessed sleep parameters among urban-dwelling youth from Mexico	Jansen et al.[66]	2021	Mexico
27	Insomnia among Japanese adolescents: a nationwide representative survey	Kaneita et al.[75]	2006	Japan
28	Associations of sleep duration and quality with disinhibited eating behaviors in adolescent girls at-risk for type 2 diabetes	Kelly et al.[53]	2016	United States
29	Cross-sectional study of randomly selected 18-year-old students showed that body mass index was only associated with sleep duration in girls	Kjartansdóttir et al.[93]	2018	Iceland
30	Sleep restriction is not associated with a positive energy balance in adolescent boys	Klingenberg et al.[54]	2012	Denmark
31	Do sleep-deprived adolescents make less-healthy food choices?	Kruger et al.[34]	2014	United States
32	Sleep duration's association with diet, physical activity, mental status, and weight among Korean high school students	Lee[51]	2017	South Korea
33	Associations of weekday and weekend sleep with children's reported eating in the absence of hunger	LeMay-Russel et al.[78]	2019	United States
34	Interactions between energy drink consumption and sleep problems: associations with alcohol use among young adolescents	Marmorstein[57]	2017	United States
35	Neural mechanisms that promote food consumption following sleep loss and social stress: an fMRI study in adolescent girls with overweight/obesity	Jensen et al.[101]	2021	United States
36	Associations of sleep duration and social jetlag with cardiometabolic risk factors in the study of Latino youth	Johnson et al.[69]	2020	United States
37	Factors associated with sleep duration among pupils	Kohyama et al.[94]	2020	Japan
38	Association between self-reported sleep duration and dietary nutrients in Korean adolescents: a population-based study	Lee et al.[65]	2020	Korea
39	Sleep-related problems and eating habits during COVID-19 lockdown in a southern Brazilian youth sample	López-Gil et al.[37]	2021	Brazil
40	Relationships between hours of sleep and health-risk behaviors in US adolescent students	McKnight-Eily et al.[48]	2011	United States
41	The association between sleep duration, sleep quality, and food consumption in adolescents: a cross-sectional study using the Korea Youth Risk Behavior Web-based Survey	Min et al.[44]	2018	Korea
42	Energy drink consumption among Australian adolescents associated with a cluster of unhealthy dietary behaviours and short sleep duration	Nuss et al.[95]	2021	Australia
43	Clustering of energy balance-related behaviours, sleep, and overweight among Finnish adolescents	Nuutinen et al.[96]	2017	Finland
44	Association between unhealthy dietary behaviors and sleep disturbances among Japanese adolescents: a nationwide representative survey	Otsuka et al.[79]	2019	Japan
45	Associations of sleep with food cravings and loss-of-control eating in youth: an ecological momentary assessment study	Parker et al.[77]	2021	United States
46	Clustering of dietary patterns, lifestyles, and overweight among Spanish children and adolescents in the ANIBES study	Perez-Rodrigo et al.[97]	2015	Spain
47	Sleep duration and consumption of sugar-sweetened beverages and energy drinks among adolescents	Sampasa-Kanyinga et al.[47]	2018	Canada
48	Association between short time in bed, health-risk behaviors and poor academic achievement among Norwegian adolescents	Stea et al.[50]	2014	Norway

Continued on the following page

TABLE 2 (continued)
Publication characteristics of the studies included from the systematic search

No.	Title	Authors	Year	Country
49	Breakfast skipping in Greek schoolchildren connected to an unhealthy lifestyle profile. Results from the National Action for Children's Health program	Tambalis et al.[71]	2019	Greece
50	Insufficient sleep duration is associated with dietary habits, screen time, and obesity in children	Tambalis et al.[98]	2018	Greece
51	Sleep duration and behavioral correlates in middle and high school students: a cross-sectional study in Zhejiang province, China	Wang et al.[38]	2021	China
52	The association of sleep duration with adolescents' fat and carbohydrate consumption	Weiss et al.[99]	2010	United States
53	Self-reported sleep duration and weight-control strategies among US high school students	Wheaton et al.[80]	2013	United States
54	Sleep duration and weight-related behaviors among adolescents	Widome et al.[36]	2019	United States

slept less than eight hours consumed a higher proportion of calories from fat and a lower proportion of calories from carbohydrates, compared to adolescents sleeping eight hours or more.[60] Another study observed that girls who slept less than five hours a night ate a higher proportion of carbohydrates.[61]

Dietary quality

Six studies assessed dietary quality.[67-72] The findings related to dietary quality were mixed, with two studies reporting that insufficient sleep was associated with poorer dietary quality[67,72] and another two studies reporting no significant relationship between sleep duration and dietary quality.[68,69] Furthermore, one study reported a significant association between sleep duration and dietary quality in the context of the association between perceived stress and dietary quality.[70]

Eating habits

Meal consumption patterns

Among the studies that examined eating habits, meal consumption pattern was the most commonly examined in relation to sleep duration. Breakfast consumption was examined in all studies except one.[59] Lunch, dinner and snack consumption was examined in a very few studies.[73,74] Among the studies examining eating habits, it was commonly found that skipping a meal, especially breakfast, was associated with less optimal sleep duration.[30,36-38,42,43,50,71] Two studies found that the prevalence of sleep disorders, such as insomnia, were higher for meal skippers.[69,75]

Eating-related cognitions

Cognitive factors associated with eating emerged as a theme in the literature. Using an experimental study design, one study found that following sleep restriction, adolescents performed more poorly on food-related inhibitory control,[76] and another found that short sleep duration was associated with loss-of-control eating.[77] Additionally, an interaction was identified whereby adolescents with a BMI in the normal weight range had a heightened response to food reward following sleep restriction. One study found that average weekday sleep duration was negatively associated with eating in the absence of hunger, and the reverse was observed for average weekend sleep duration.[78]

Eating contexts

Contextual factors surrounding eating were identified as a theme in the variables investigated. One study measured a number of weight-related behaviours, including eating habits such as eating when full.[36] Another study found that sleep duration provided a partial explanation

TABLE 3
Publication characteristics of the studies included from the grey literature review

No.	Title	Author	Year	Country
1	#consumingitall: Understanding the complex relationship between media consumption and eating behaviors	Albert[41]	2017	United States
2	Are eating habits associated with adequate sleep among high school students?	Bhurosy and Thiagarajah[30]	2020	United States
3	Obesity and sleep: assessing risk among African American adolescent girls in Chicago	Brakefield[100]	2012	United States
4	European adolescents' level of perceived stress is inversely related to their diet quality: the Healthy Lifestyle in Europe by Nutrition in Adolescence study	De Vriendt et al.[70]	2012	Belgium
5	Sleep quality and duration is related with diet and obesity in young adolescent living in Sicily, Southern Italy	Ferranti et al.[32]	2016	Italy
6	Characteristics associated with sleep duration, chronotype, and social jet lag in adolescents	Malone et al.[40]	2016	United States
7	Association of sleep duration and snack consumption in children and adolescents: the CASPIAN- V study	Mozaffarian et al.[35]	2020	Iran

TABLE 4
Study characteristics of included studies

Characteristic		Frequency	%
Year	2006–2010	2	0.03
	2011–2015	15	24.6
	2016–2020	39	63.9
	2021–2025	5	8.2
Region	Asia	8	13.1
	Europe	14	23.0
	Middle East	7	11.5
	North America	22	36.1
	Oceania	3	4.9
	South America	7	11.5
Publication type	Dissertation	2	3.3
	Original research	59	96.7
Study design	Cross-over	2	3.3
	Cross-sectional	53	86.9
	Prospective	5	8.2
	Randomized control	1	1.6
Study type	Experimental	3	4.9
	Intervention	1	1.6
	Observational	57	93.4
Setting	School	35	57.4
	Clinic	12	19.7
	Combination[a]	1	1.6
	Community	12	19.7
	Laboratory	1	1.6
Age group	High school–aged and younger	31	50.8
	High school–aged and older	2	3.3
	High school–aged only	25	41.0
	Not specified	3	4.9
Sex/gender	Female only	3	4.9
	Male only	1	1.6
	Both	54	88.5
	Not specified	3	4.9

[a] One study used a sample from a sport club and a school.

for the relationship between media consumption (e.g. listening to music, watching television, playing video games, instant messaging, emailing) and eating behaviours, but only for specific media and only in males.[41] One study found that longer sleep duration was associated with fewer times eating outside of the home in a week.[32] The findings from a study conducted in Japan indicated that short sleep duration was associated with family meal frequency.[79]

Disordered eating

Two studies found that very short sleep duration was significantly associated with weight control strategies, such as fasting and purging and eating fewer calories, among adolescent boys and girls.[38,80] Another study examined the associations among sleep duration, daytime sleepiness and disinhibited eating, including binge eating.[53] One study observed that emotional and night eating emerged during the COVID-19 pandemic, and was associated with symptoms of insomnia.[59]

Discussion

Summary of evidence

The objective of this scoping review was to explore and synthesize the literature on sleep duration and eating behaviours in adolescents. In total, 61 articles were included in this synthesis. This review also mapped out the characteristics of existing research by examining the research designs, study populations, outcome variables and research gaps in the current body of literature. To our knowledge, this is the first synthesis published on the topic.

The majority of studies were observational, employed a cross-sectional design, used a school-based population and were published in North America. With respect to methodologies used when assessing eating behaviours, the review identified that the current research focussed heavily on the intake of food and beverages through self-reported questionnaires. This synthesis also revealed a heavy emphasis on eating behaviours related to food group intakes, such as fruit and vegetables. Intakes of beverages and processed food were also very commonly investigated variables. Surprisingly, few studies examined eating habits in relation to sleep duration. Among the studies that examined eating habits, the majority focussed on breakfast consumption, with very few studies including measures of eating-related cognitions, eating contexts or disordered eating.

A prominent gap in the literature is the limited examination of eating habits as opposed to food consumption. Only four studies examined the contextual factors that surround eating, including eating in the absence of hunger, eating with family and friends and eating while consuming media, in relation to sleep duration in adolescents. This gap is critical to address because of the influence of sleep on eating behaviours, and the increasing recognition of the role of eating habits in overall healthy eating practices.[3,81] Therefore, further investigation into the connections between sleep duration and eating habits of adolescents is warranted.

An area that remains to be addressed is how sleep duration is implicated in disordered eating behaviours among adolescents. Previous research has identified that disordered eating and eating disorders often emerge during adolescence and early adulthood.[82] However, there is limited research examining the association between sleep duration and disordered eating among adolescents. Of particular

TABLE 5
Research focus of sleep duration and eating behaviour research

Eating behaviour variable	n
Food group intake (e.g. vegetables, fruit, meat, milk)	25
Beverage intake (e.g. sugar-containing beverages)	22
Meal consumption patterns (e.g. skipping breakfast)	18
Processed food intake (e.g. chocolate, candy, fried potatoes)	15
Energy intake (e.g. total calories)	11
Caffeinated beverage intake (e.g. coffee, energy drinks)	9
Eating-related cognition (e.g. dietary restraint)	7
Macronutrient intake (e.g. protein, fat, carbohydrate)	7
Dietary quality	6
Eating context (e.g. location, people)	5
Disordered eating (e.g. bingeing)	4
Eating behaviour measure	**n**
Self-reported questionnaire	44
Interview recall	9
Objective measure	4
Experimental task	2
Combination	1
Food record	1

relevance to this review is the role of insufficient sleep on binge eating. Research demonstrates that inadequate sleep is associated with binge eating, partly due to decreasing leptin (reduces appetite) and increasing ghrelin (stimulates appetite).[83] However, chronic energy restriction has also been demonstrated to compromise sleep health through mechanisms such as reducing orexin, which plays a role in regulating arousal, hunger and wakefulness,[84] and compromising sleep by increasing wake time and shallow sleep.[85] Considering the bidirectional nature of the relationship between sleep and disordered eating, this area of research requires further investigation.

Additionally, two factors that impact sleep duration and eating behaviours that were not adequately addressed are the role of stress, and changes in metabolic hormones. Research demonstrates that stressful life events impact sleep through alterations to the duration and quality of sleep; however, very few studies reviewed in this synthesis examined the influence of stress on the eating behaviours of adolescents. Research also demonstrates that those experiencing shorter sleep and stress exhibit changes in metabolic hormones (e.g. reduced leptin and elevated ghrelin), which likely contributes to an increase in appetite and changes in eating behaviours[8,86] and altered inhibitory control.[76] Therefore, addressing this gap in the literature is crucial to better understanding the potential moderating effects of stress on sleep and eating-related cognitions, such as eating in the absence of hunger and disinhibited eating.

The overwhelming majority of studies published on this topic used a cross-sectional approach and were observational in nature. Self-report measures were used the majority of the time as indicators of sleep duration and eating behaviours. To develop a clearer understanding of the nature of the relationship between sleep duration and eating behaviours in adolescents, a wider variety of research designs and methods should be employed. The predominantly cross-sectional nature of the study designs and analyses does not enable inferences into the temporality of the associations observed. Future studies using a prospective cohort design are required to assess the temporality and bidirectional nature of the associations between sleep duration and eating behaviours.

Strengths and limitations

One of the strengths of this review is that we included the optional step of engaging stakeholders.[23] By engaging stakeholders, we were able to share and validate preliminary findings and solicit the perspectives of researchers and clinicians working in the community. Additionally, reliability checks were conducted throughout the scoping review, incorporating steps such as having two or more members of the research team review articles during the data selection and extraction stages.

There are limitations to this review. Unlike other kinds of reviews (e.g. systematic, meta-analytical), scoping reviews are not designed to evaluate the strength of associations between the variables or the quality of studies that were reviewed.[87,88] Thus, neither the strength of the associations observed nor the quality of the included studies in our review were assessed. Instead, scoping reviews are used to gather information from a range of study designs and methods in order to identify the types of evidence available in a given field and to identify knowledge gaps; they can serve as a precursor to a systematic review.[89]

Another limitation to this review is that only studies published in English were screened.

Finally, there is a possibility that relevant articles were inadvertently excluded. Although the search strategy was designed in consultation with subject specialist librarians, less commonly used terms in the literature may have been overlooked in the final search strategy.

Conclusion

Although research on sleep duration and eating behaviours in adolescent populations has been increasingly published in the past decade, much remains to be examined in this field. Further research on this topic is necessary in order to better understand how ensuring sufficient sleep among adolescents can support healthier eating practices. Future research should investigate how insufficient sleep may impact the eating habits of adolescents, including eating-related cognition, eating contexts and disordered eating behaviours. These lines of inquiry could contribute to supporting healthy eating among adolescents and informing behavioural interventions aimed at managing diet-related conditions.

Acknowledgements

The authors would like to thank the liaison librarians at the University of Waterloo

for their guidance in developing the literature search strategy and review protocol.

Conflicts of interest

None.

Funding

This research did not receive any specific grants from funding agencies in the public, commercial or not-for-profit sectors.

Authors' contributions and statement

All authors contributed meaningfully to the preparation, drafting and editing of this paper. ND designed the protocol and led all aspects of the study, including data collection, extraction, charting, synthesis, stakeholder consultations and writing. AP, KR and EVB engaged in data collection and extraction. KR wrote the introduction, ND wrote the methods and results and AP wrote the discussion. EVB critically reviewed all components of the manuscript. MAF supervised the research, revised the manuscript and approved the final 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.

References

1. Bacon SL, Campbell NRC, Raine KD, et al. Canada's new Healthy Eating Strategy: implications for health care professionals and a call to action. Can Pharm J. 2019;152(3):151-157. https://doi.org/10.1177/1715163519834891

2. Te Morenga L, Montez JM. Health effects of saturated and trans-fatty acid intake in children and adolescents: systematic review and meta-analysis. PLoS ONE. 2017;12(11): e0186672. https://doi.org/10.1371/JOURNAL.PONE.0186672

3. Health Canada. Canada's food guide [Internet]. Ottawa (ON): Government of Canada; 2019 [cited 2019 May]. Available from: https://food-guide.canada.ca/en/

4. Haerens L, Craeynest M, Deforche B, Maes L, Cardon G, De Bourdeaudhuij I. The contribution of psychosocial and home environmental factors in explaining eating behaviours in adolescents. Eur J Clin Nutr. 2008;62(1): 51-9. https://doi.org/10.1038/sj.ejcn.1602681

5. Córdova F, Barja S, Brockmann P. Consequences of short sleep duration on the dietary intake in children: a systematic review and metanalysis. Sleep Med Rev. 2018;42:68-84. https://doi.org/10.1016/J.SMRV.2018.05.006

6. Capers P, Fobian A, Kaiser K, Borah R, Allison D. A systematic review and meta-analysis of randomized controlled trials of the impact of sleep duration on adiposity and components of energy balance. Obes Rev. 2015;16(9):771-82. https://doi.org/10.1111/OBR.12296

7. Grandner MA. Sleep and obesity risk in adults: possible mechanisms; contextual factors; and implications for research, intervention, and policy. Sleep Health. 2017;3(5):393-400. https://doi.org/10.1016/J.SLEH.2017.07.014

8. Taheri S, Lin L, Austin D, Young T, Mignot E. Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index. PLoS Med. 2004;1(3):e62. https://doi.org/10.1371/JOURNAL.PMED.0010062

9. Burrows T, Fenton S, Duncan M. Diet and sleep health: a scoping review of intervention studies in adults. J Hum Nutr Diet. 2020;33(3):308-29. https://doi.org/10.1111/JHN.12709

10. Fenton S, Burrows T, Skinner J, Duncan M. The influence of sleep health on dietary intake: a systematic review and meta-analysis of intervention studies. J Hum Nutr Diet. 2021; 34(2):273-85. https://doi.org/10.1111/JHN.12813

11. Chaput J-P. Sleep patterns, diet quality and energy balance. Physiol Behav. 2014;134:86-91. https://doi.org/10.1016/J.PHYSBEH.2013.09.006

12. World Cancer Research Fund, American Institute for Cancer Research. Diet, nutrition, physical activity and cancer: a global perspective. [Continuous Update Project Expert Report]. World Cancer Research Fund; 2018. 116 p. Available from: https://www.wcrf.org/wp-content/uploads/2021/02/Summary-of-Third-Expert-Report-2018.pdf

13. Felső R, Lohner S, Hollódy K, Erhardt É, Molnár D. Relationship between sleep duration and childhood obesity: systematic review including the potential underlying mechanisms. Nutr Metab Cardiovasc Dis. 2017;27(9):751-61. https://doi.org/10.1016/j.numecd.2017.07.008

14. Crowley SJ, Acebo C, Carskadon MA. Sleep, circadian rhythms, and delayed phase in adolescence. 2007;8(6):602-12. https://doi.org/10.1016/j.sleep.2006.12.002

15. Chaput J-P, Gray CE, Poitras VJ, et al. Systematic review of the relationships between sleep duration and health indicators in school-aged children and youth. Appl Physiol Nutr Metab. 2016;41(6 Suppl 3):S266-S282. https://doi.org/10.1139/APNM-2015-0627

16. Patte KA, Qian W, Leatherdale ST. Sleep duration trends and trajectories among youth in the COMPASS study. Sleep Health. 2017;3(5):309-16. https://doi.org/10.1016/j.sleh.2017.06.006

17. Roberts RE, Roberts CR, Xing Y. Restricted sleep among adolescents: prevalence, incidence, persistence, and associated factors. Behav Sleep Med. 2011;9(1):18-30. https://doi.org/10.1080/15402002.2011.533991

18. Canadian Society for Exercise Physiology (CSEP). Children (5–11 years) and youth (12–17 years) [Internet]. Ottawa (ON): CSEP; 2016 [cited 2021 Aug 26]. Available from: https://csepguidelines.ca/guidelines/children-youth

19. Owens J, Adolescent Sleep Working Group, Committee on Adolescence. Insufficient sleep in adolescents and young adults: an update on causes and consequences. Pediatrics. 2014; 134(3):e921-e932. https://doi.org/10.1542/PEDS.2014-1696

20. Beebe DW, Simon S, Summer S, Hemmer S, Strotman D, Dolan LM. Dietary intake following experimentally restricted sleep in adolescents. Sleep. 2013;36(6):827-34. https://doi.org/10.5665/SLEEP.2704

21. Godos J, Grosso G, Castellano S, Galvano F, Caraci F, Ferri R. Association between diet and sleep quality: a systematic review. Sleep Med Rev. 2021;57. https://doi.org/10.1016/J.SMRV.2021.101430

22. Dashti H, Scheer F, Jacques P, Lamon-Fava S, Ordovás J. Short sleep duration and dietary intake: epidemiologic evidence, mechanisms, and health implications. Adv Nutr. 2015;6(6):648-59. https://doi.org/10.3945/an.115.008623

23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 2005; 8(1):19-32. https://doi.org/10.1080/1364557032000119616

24. Levac D, Colquhoun H, O'Brien K. Scoping studies: advancing the methodology. Implement Sci. 2010;5:69. https://doi.org/10.1186/1748-5908-5-69

25. Doan N, Ferro MA. Sleep duration and eating behaviours in youth: a scoping review protocol. BMJ Open. 2019;9(12):e030457. https://doi.org/10.1136/BMJOPEN-2019-030457

26. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7): 467-73. https://doi.org/10.7326/M18-0850

27. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535. https://doi.org/10.1136/bmj.b2535

28. Godin K, Stapleton J, Kirkpatrick SI, Hanning RM, Leatherdale ST. Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. Syst Rev. 2015;4:138. https://doi.org/10.1186/s13643-015-0125-0

29. Daudt HM, van Mossel C, Scott SJ. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol. 2013;13:48. https://doi.org/10.1186/1471-2288-13-48

30. Bhurosy T, Thiagarajah K. Are eating habits associated with adequate sleep among high school students? J Sch Health. 2020;90(2):81-7. https://doi.org/10.1111/JOSH.12852

31. Cao M, Zhu Y, Sun F, Luo J, Jing J. Short sleep duration is associated with specific food intake increase among school-aged children in China: a national cross-sectional study. BMC Public Health. 2019;19(1):558. https://doi.org/10.1186/S12889-019-6739-8

32. Ferranti R, Marventano S, Castellano S, et al. Sleep quality and duration is related with diet and obesity in young adolescent living in Sicily, Southern Italy. Sleep Sci. 2016;9(2):117-22. https://doi.org/10.1016/J.SLSCI.2016.04.003

33. Garaulet M, Ortega FB, Ruiz JR, et al. Short sleep duration is associated with increased obesity markers in European adolescents: effect of physical activity and dietary habits. the HELENA study. Int J Obes. 2011; 35(10):1308-17. https://doi.org/10.1038/IJO.2011.149

34. Kruger AK, Reither EN, Peppard PE, Krueger PM, Hale L. Do sleep-deprived adolescents make less-healthy food choices? Br J Nutr. 2014;111(10):1898-904. https://doi.org/10.1017/S0007114514000130

35. Mozaffarian N, Heshmat R, Ataie-Jafari A, et al. Association of sleep duration and snack consumption in children and adolescents: the CASPIAN-V study. Food Sci Nutr. 2020;8(4):1888-97. https://doi.org/10.1002/FSN3.1471

36. Widome R, Lenk KM, Laska MN, et al. Sleep duration and weight-related behaviors among adolescents. Child Obes. 2019;15(7):434-42. https://doi.org/10.1089/CHI.2018.0362

37. López-Gil JF, Gaya AR, Reuter CP, et al. Sleep-related problems and eating habits during COVID-19 lockdown in a southern Brazilian youth sample. Sleep Med. 2021;85:150-6. https://doi.org/10.1016/J.SLEEP.2021.07.003

38. Wang M, Zhong JM, Hu RY, Gong WW, Yu M. Sleep duration and behavioral correlates in middle and high school students: a cross-sectional study in Zhejiang province, China. Sleep Med. 2021;79:55-61. https://doi.org/10.1016/J.SLEEP.2021.01.008

39. Garcez MR, de Castro MA, César CL, Goldbaum M, Fisberg RM. A chrononutrition perspective of diet quality and eating behaviors of Brazilian adolescents in associated with sleep duration. Chronobiol Int. 2021;38(3): 387-99. https://doi.org/10.1080/07420528.2020.1851704

40. Malone SK, Zemel B, Compher C, et al. Characteristics associated with sleep duration, chronotype, and social jet lag in adolescents. J Sch Nurs. 2016; 32(2):120-31. https://doi.org/10.1177/1059840515603454

41. Albert SL. #consumingitall: Understanding the complex relationship between media consumption and eating behaviors [dissertation]. [Los Angeles]: University of California; 2017. 200 p. Available from: https://www.proquest.com/docview/1914676841?pq-origsite=gscholar&fromopenview=true

42. Al-Haifi AA, AlMajed HT, Al-Hazzaa HM, Musaiger AO, Arab MA, and Hasan RA. Relative contributions of obesity, sedentary behaviors and dietary habits to sleep duration among Kuwaiti adolescents. Global J Health Sci. 2016;8(1):107-17. https://doi.org/10.5539/gjhs.v8n1p107

43. Al-Hazzaa HM, Musaiger AO, Abahussain NA, Al-Sobayel HI, Qahwaji DM. Lifestyle correlates of self-reported sleep duration among Saudi adolescents: a multicentre school-based cross-sectional study. Child Care Health Dev. 2014; 40(4):533-42. https://doi.org/10.1111/CCH.12051

44. Min C, Kim HJ, Park IS, et al. The association between sleep duration, sleep quality, and food consumption in adolescents: a cross-sectional study using the Korea Youth Risk Behavior Web-based Survey. BMJ Open. 2018; 8(7):e022848. https://doi.org/10.1136/BMJOPEN-2018-022848

45. Jansen EC, Baylin A, Cantoral A, et al. Dietary patterns in relation to prospective sleep duration and timing among Mexico City adolescents. Nutrients. 2020;12(8):2305. https://doi.org/10.3390/NU12082305

46. André B, Canhão H, Espnes GA, et al. Association between food patterns and difficulties in falling asleep among adolescents in Norway—a descriptive Young-Hunt3 study. J Public Health. 2021;29(6):1373-8. https://doi.org/10.1007/S10389-020-01249-4

47. Sampasa-Kanyinga H, Hamilton HA, Chaput J-P. Sleep duration and consumption of sugar-sweetened beverages and energy drinks among adolescents. Nutrition. 2018;48:77-81. https://doi.org/10.1016/j.nut.2017.11.013

48. McKnight-Eily LR, Eaton DK, Lowry R, Croft JB, Presley-Cantrell L, Perry GS. Relationships between hours of sleep and health-risk behaviors in US adolescent students. Prev Med (Baltim). 2011;53(4-5):271-3. https://doi.org/10.1016/J.YPMED.2011.06.020

49. Almulla AA, Zoubeidi T. Association of overweight, obesity and insufficient sleep duration and related lifestyle factors among school children and adolescents. Int J Adolesc Med Health. 2021;34(2):31-40. https://doi.org/10.1515/ijamh-2021-0041

50. Stea TH, Knutsen T, Torstveit MK. Association between short time in bed, health-risk behaviors and poor academic achievement among Norwegian adolescents. Sleep Med. 2014;15(6): 666-71. https://doi.org/10.1016/J.SLEEP.2014.01.019

51. Lee J. Sleep duration's association with diet, physical activity, mental status, and weight among Korean high school students. Asia Pac J Clin Nutr. 2017;26(5):906-13. https://doi.org/10.6133/apjcn.082016.04

52. Golley RK, Maher CA, Matricciani L, Olds TS. Sleep duration or bedtime? Exploring the association between sleep timing behaviour, diet and BMI in children and adolescents. Int J Obes. 2013;37(4):546-51. https://doi.org/10.1038/IJO.2012.212

53. Kelly NR, Shomaker LB, Radin RM, et al. Associations of sleep duration and quality with disinhibited eating behaviors in adolescent girls at-risk for type 2 diabetes. Eat Behav. 2016;22: 149-55. https://doi.org/10.1016/J.EATBEH.2016.06.019

54. Klingenberg L, Chaput J-P, Holmbäck U, Jennum P, Astrup A, Sjödin A. Sleep restriction is not associated with a positive energy balance in adolescent boys. Am J Clin Nutr. 2012;96(2):240-8. https://doi.org/10.3945/AJCN.112.038638

55. Hayes JF, Balantekin KN, Altman M, Wilfley DE, Taylor BC, Williams J. Sleep patterns and quality are associated with severity of obesity and weight-related behaviors in adolescents with overweight and obesity. Child Obes. 2018;14(1):11-17. https://doi.org/10.1089/CHI.2017.0148

56. He F, Bixler EO, Berg A, et al. Habitual sleep variability, not sleep duration, is associated with caloric intake in adolescents. Sleep Med. 2015;16(7): 856-61. https://doi.org/10.1016/J.SLEEP.2015.03.004

57. Marmorstein NR. Interactions between energy drink consumption and sleep problems: associations with alcohol use among young adolescents. J Caffeine Res. 2017;7(3):111-6. https://doi.org/10.1089/JCR.2017.0007

58. Galland BC, de Wilde T, Taylor RW, Smith C. Sleep and pre-bedtime activities in New Zealand adolescents: differences by ethnicity. Sleep Health. 2020;6(1):23-31. https://doi.org/10.1016/J.SLEH.2019.09.002

59. Alamrawy R, Fadl N, Khaled A. Psychiatric morbidity and dietary habits during COVID-19 pandemic: a cross-sectional study among Egyptian Youth (14–24 years). Middle East Curr Psychiatry. 2021;28(1):6. https://doi.org/10.1186/s43045-021-00085-w

60. Weiss R, Stumbo PJ, Divakaran A. Automatic food documentation and volume computation using digital imaging and electronic transmission. J Am Diet Assoc. 2010;110(1):42-4. https://doi.org/10.1016/j.jada.2009.10.011

61. Al-Disi D, Al-Daghri N, Khanam L, et al. Subjective sleep duration and quality influence diet composition and circulating adipocytokines and ghrelin levels in teen-age girls. Endocr J. 2010;57(10):915-23. https://doi.org/10.1507/ENDOCRJ.K10E-145

62. Beebe DW, Miller N, Kirk S, Daniels SR, Amin R. The association between obstructive sleep apnea and dietary choices among obese individuals during middle to late childhood. Sleep Med. 2011;12(8):797-9. https://doi.org/10.1016/j.sleep.2010.12.020

63. Ievers-Landis CE, Kneifel A, Giesel J, et al. Dietary intake and eating-related cognitions related to sleep among adolescents who are overweight or obese. J Ped Psychol. 2016;41(6):670-9. https://doi.org/10.1093/jpepsy/jsw017

64. He F, Dong H, Fernandez-Mendoza J, Bixler EO, Liao J, Liao D. Racial/ethnic disparity in habitual sleep is modified by caloric intake in adolescents. Sleep Med. 2020;76:65-71. https://doi.org/10.1016/j.sleep.2020.10.001

65. Lee JH, Chung SJ, Seo WH. Association between self-reported sleep duration and dietary nutrients in Korean adolescents: a population-based study. Children. 2020;7(11):221. https://doi.org/10.3390/children7110221

66. Jansen EC, Corcoran K, Perng W, et al. Relationships of beverage consumption and actigraphy-assessed sleep parameters among urban-dwelling youth from Mexico. Public Health Nutr. 2021:1-10. https://doi.org/10.1017/S136898002100313X

67. Bel S, Michels N, De Vriendt T, et al. Association between self-reported sleep duration and dietary quality in European adolescents. Br J Nutr. 2013;110(5):949-59. https://doi.org/10.1017/S0007114512006046

68. Adelantado-Renau M, Beltran-Valls MR, Esteban-Cornejo I, Martínez-Vizcaíno V, Santaliestra-Pasías AM, Moliner-Urdiales D. The influence of adherence to the Mediterranean diet on academic performance is mediated by sleep quality in adolescents. Acta Paediatr. 2019;108(2):339-46. https://doi.org/10.1111/APA.14472

69. Johnson DA, Reid M, Vu TH, et al. Associations of sleep duration and social jetlag with cardiometabolic risk factors in the study of Latino youth. Sleep Health. 2020;6(5):563-9. https://doi.org/10.1016/J.SLEH.2020.02.017

70. De Vriendt T, Clays E, Huybrechts I, et al. European adolescents' level of perceived stress is inversely related to their diet quality: The Healthy Lifestyle in Europe by Nutrition in Adolescence study. Br J Nutr. 2012;108(2):371-80. https://doi.org/10.1017/S0007114511005708

71. Tambalis KD, Panagiotakos DB, Psarra G, Sidossis LS. Breakfast skipping in Greek schoolchildren connected to an unhealthy lifestyle profile. Results from the National Action for Children's Health program. Nutr Diet. 2019;76(3):328-35. https://doi.org/10.1111/1747-0080.12522

72. Garcez MR, de Castro MA, César CL, Goldbaum M, Fisberg RM. A chrononutrition perspective of diet quality and eating behaviors of Brazilian adolescents in associated with sleep duration. Chronobiol Int. 2021;38(3):387-99. https://doi.org/10.1080/07420528.2020.1851704

73. Haghighatdoost F, Kelishadi R, Qorbani M, et al. Family dinner frequency is inversely related to mental disorders and obesity in adolescents: the CASPIAN-III study. Arch Iran Med. 2017;20(4):218-23.

74. Hoefelmann LP, Silva KS, Filho VC, da Silva JA, Nahas MV. Behaviors associated to sleep among high school students: cross-sectional and prospective analysis. Rev Bras Cineantropometria Desempenho Hum. 2014;16(Suppl. 1):68-78. https://doi.org/10.5007/1980-0037.2014V16S1P68

75. Kaneita Y, Ohida T, Osaki Y, et al. Insomnia among Japanese adolescents: a nationwide representative survey. Sleep. 2006;29(12):1543-50. https://doi.org/10.1093/SLEEP/29.12.1543

76. Duraccio KM, Zaugg K, Jensen CD. Effects of sleep restriction on food-related inhibitory control and reward in adolescents. J Pediatr Psychol. 2019;44(6):692-702. https://doi.org/10.1093/JPEPSY/JSZ008

77. Parker MN, LeMay-Russell S, Schvey NA, et al. Associations of sleep with food cravings and loss-of-control eating in youth: an ecological momentary assessment study. Pediatr Obes. 2021;17(2):e12851. https://doi.org/10.1111/IJPO.12851

78. LeMay-Russell S, Tanofsky-Kraff M, Schvey NA, et al. Associations of weekday and weekend sleep with children's reported eating in the absence of hunger. Nutrients. 2019;11(7):1658. https://doi.org/10.3390/nu11071658

79. Otsuka Y, Kaneita Y, Itani O, et al. Association between unhealthy dietary behaviors and sleep disturbances among Japanese adolescents: a nationwide representative survey. Sleep Biol Rhythms. 2019;17(1):93-102. https://doi.org/10.1007/s41105-018-0193-3

80. Wheaton AG, Perry GS, Chapman DP, Croft JB. Self-reported sleep duration and weight-control strategies among US high school students. Sleep. 2013;36(8):1139-45. https://doi.org/10.5665/SLEEP.2872

81. Ministry of Health of Brazil. Dietary guidelines for the Brazilian population. 2nd ed. [Internet]. Pan American Health Organization; 2014 [cited 2021 Aug 27]. Available from: https://www3.paho.org/hq/index.php?option=com_content&view=article&id=11564:dietary-guidelines-brazilian-population&Itemid=4256&lang=en

82. Hudson JI, Hiripi E, Pope HG Jr, Kessler RC. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. Biol Psychiatry. 2007;61(3):348-58. https://doi.org/10.1016/J.BIOPSYCH.2006.03.040

83. Trace SE, Thornton LM, Runfola CD, Lichtenstein P, Pedersen NL, Bulik CM. Sleep problems are associated with binge eating in women. Int J Eat Disord. 2012;45(5):695-703. https://doi.org/10.1002/EAT.22003

84. Davis JF, Choi DL, Benoit SC. Orexigenic hypothalamic peptides behavior and feeding. In: Preedy V, Watson R, Martin C, editors, Handbook of Behavior, Food and Nutrition [e-book]. New York (NY): Springer; 2011:355-369. https://doi.org/10.1007/978-0-387-92271-3_24

85. Lauer CJ, Krieg J-C. Sleep in eating disorders. Sleep Med Rev. 2004;8(2):109-18. https://doi.org/10.1016/S1087-0792(02)00122-3

86. Macedo DM, Diez-Garcia RW. Sweet craving and ghrelin and leptin levels in women during stress. Appetite. 2014;80:264-70. https://doi.org/10.1016/J.APPET.2014.05.031

87. Sucharew H, Macaluso M. Methods for research evidence synthesis: the scoping review approach. J Hosp Med. 2019;14(7):416-18. https://doi.org/10.12788/JHM.3248

88. Green BN, Johnson CD, Adams A. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. J Chiropr Med. 2006;5(3):101-17. https://doi.org/10.1016/S0899-3467(07)60142-6

89. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol. 2018;18:143. https://doi.org/10.1186/S12874-018-0611-X

90. Collese TS, De Moraes AC, Fernández-Alvira JM, et al. How do energy balance-related behaviors cluster in adolescents? Int J Public Health. 2019;64(2):195-208. https://doi.org/10.1007/s00038-018-1178-3

91. Duraccio KM, Krietsch KN, Zhang N, et al. The impact of short sleep on food reward processes in adolescents. J Sleep Res. 2021;30(2):e13054. https://doi.org/10.1111/jsr.13054

92. Hoefelmann LP, Silva KS, Lopes AD, et al. Association between unhealthy behavior and sleep quality and duration in adolescents. Rev Bras Cineantropom Desempenho Hum. 2015; 17:318-27. https://doi.org/10.5007/1980-0037.2015v17n3p318

93. Kjartansdottir I, Arngrimsson SA, Bjarnason R, Olafsdottir AS. Cross-sectional study of randomly selected 18-year-old students showed that body mass index was only associated with sleep duration in girls. Acta Paediatr. 2018;107(6):1070-6. https://doi.org/10.1111/apa.14238

94. Kohyama J, Ono M, Anzai Y. Factors associated with sleep duration among pupils. Ped Int. 2020;62(6):716-24. https://doi.org/10.1111/PED.14178

95. Nuss T, Morley B, Scully M, Wakefield M. Energy drink consumption among Australian adolescents associated with a cluster of unhealthy dietary behaviours and short sleep duration. Nutr J. 2021;20:64. https://doi.org/10.1186/s12937-021-00719-z

96. Nuutinen T, Lehto E, Ray C, Roos E, Villberg J, Tynjälä J. Clustering of energy balance-related behaviours, sleep, and overweight among Finnish adolescents. Int J Public Health. 2017; 62(8):929-38. https://doi.org/10.1007/s00038-017-0991-4

97. Pérez-Rodrigo C, Gil Á, González-Gross M, et al. Clustering of dietary patterns, lifestyles, and overweight among Spanish children and adolescents in the ANIBES study. Nutrients. 2015;8(1):11. https://doi.org/10.3390/nu8010011

98. Tambalis KD, Panagiotakos DB, Psarra G, Sidossis LS. Insufficient sleep duration is associated with dietary habits, screen time, and obesity in children. J Clinical Sleep Med. 2018; 14(10):1689-96. https://doi.org/10.5664/jcsm.7374

99. Weiss A, Xu F, Storfer-Isser A, Thomas A, Ievers-Landis CE, Redline S. The association of sleep duration with adolescents' fat and carbohydrate consumption. Sleep. 2010;33(9):1201-9. https://doi.org/10.1093/sleep/33.9.1201

100. Brakefield T. Obesity and sleep: assessing risk among African American adolescent girls in Chicago [dissertation]. [Chicago]: Rosalind Franklin University of Medicine and Science; 2012. 75 p.

101. Jensen CD, Zaugg KK, Muncy NM, et al. Neural mechanisms that promote food consumption following sleep loss and social stress: an fMRI study in adolescent girls with overweight/obesity. Sleep. 2022;45(3):zsab263. https://doi.org/10.1093/sleep/zsab263

https://doi.org/10.24095/hpcdp.42.9.03

Original quantitative research

Correlates of perceived success of health-promoting interventions in elementary schools

Erin K. O'Loughlin, PhD (1,2); Jodi Kalubi, BSc(N) (2,3,4)*; Teodora Riglea, MSc (2); Annie Pelekanakis, MSc (2); Jennifer L. O'Loughlin, PhD (2,3)*

This article has been peer reviewed.

Abstract

Introduction: School-based health-promoting interventions (HPIs) aim to support youth development and positively influence modifiable lifestyle behaviours. Identifying factors that contribute to or hinder the perceived success of HPIs could facilitate their adaptation, improve implementation and contribute to HPI sustainability. The objective of this study was to identify factors in three domains (school characteristics, characteristics of the HPI and factors related to planning and implementing the HPI) associated with perceived success of HPIs among school principals in elementary schools.

Methods: Data were drawn from Project Prome*SS*, a cross-sectional survey of school principals and/or nominated staff members in a convenience sample of 171 public elementary schools in Quebec, Canada. School board and school recruitment spanned three academic school years (2016–2019). Data on school and participant characteristics, HPI characteristics, variables related to HPI planning and implementation and perceived success of the HPI were collected in two-part, structured telephone interviews. Descriptive statistics were used to characterize schools and study participants. Twenty-eight potential correlates of perceived HPI success were investigated separately in multivariable linear regression modelling.

Results: Participants generally perceived HPIs as highly successful. After controlling for number of students, language of instruction, school neighbourhood and school deprivation, we identified five correlates of perceived success, including lower teacher turnover, higher scores for school physical environment, school/teacher commitment to student health, principal leadership and school being a developer (vs. adopter) of the HPI.

Conclusion: If replicated, these factors should be considered by HPI developers and school personnel when planning and implementing HPIs in elementary schools.

Keywords: *health-promoting schools, interventions, cross-sectional study, perceived success*

Highlights

- Participants generally perceived health-promoting interventions (HPIs) as highly successful.
- Four of 11 school characteristics were associated with perceived success of the HPI, including lower teacher turnover, school physical environment, school/teacher commitment to student health and principal leadership.
- None of the eight characteristics of the HPI was associated with perceived success.
- Of the nine factors related to HPI planning or implementation, only being a developer (vs. an adopter) of the HPI was associated with perceived success.

Introduction

School-based health-promoting interventions (HPIs) support the development of positive physical, emotional and mental health among youth, including the acquisition of healthy lifestyle behaviours.[1,2] Common HPI theme areas include physical activity (which generally declines from childhood into adolescence and young adulthood[3,4]), healthy nutrition (e.g. attaining adequate levels of consumption of vegetables, fruits and whole grains), substance use behaviours (including use of alcohol, tobacco and cannabis, which can emerge early and escalate during adolescence)[5] and awareness related to aggressive behaviour (including verbal, physical and cyber-bullying). School-based HPIs are important components of broader public health strategies that aim to foster health-promoting behaviours in children from an early age.[6] Because children spend many hours each day at school, elementary schools are ideal settings for HPIs because they have high potential for reaching all children, regardless of socioeconomic status.[7]

Numerous HPIs are deemed centrally important for child development and are therefore government-mandated. In Quebec, Canada, these comprise HPIs that aim to

Author references:

1. Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, Ontario, Canada
2. Carrefour de l'innovation et de l'évaluation en santé, Centre de recherche du Centre hospitalier de l'Université de Montréal, Montréal, Québec, Canada
3. Département de médecine sociale et préventive, Université de Montréal, Montréal, Québec, Canada
4. Centre de recherche en santé publique (CReSP), Université de Montréal and CIUSSS du Centre-Sud-de-l'Île-de-Montréal, Montréal, Québec, Canada

* Co-first authors.

Correspondence: Jennifer O'Loughlin, Department of Social and Preventive Medicine, University of Montréal, 850 rue Saint-Denis, Bureau S02-370, Montréal, QC H2X 0A9; Tel: 514-890-8000 x 15858; Fax: 514-412-7137; E-mail: jennifer.oloughlin@umontreal.ca

improve awareness, knowledge and attitudes, and promote healthy behaviours related to physical activity, tobacco use, dental health, sex education and bullying.[8-11] In addition to government-mandated HPIs, many schools choose to implement other HPIs, depending on perceived need within the school community. These school-specific HPIs may be adopted by the school from an external organization or developed *de novo* by the school.

Despite its importance, evaluation of the impact of school-based HPIs can pose major challenges in assessing benefits. Many school-based HPIs are not evaluated[12,13] and among those that are evaluated, results on effectiveness, implementation success and sustainability are often mixed.[14] For example, two reviews[12,15] suggested small to modest effects of tobacco prevention programs, although evidence on long-term effects is limited and most studies are of relatively poor quality.[12] Methodological challenges include ethical and feasibility issues in implementing randomized controlled trials (RCTs) in a school context[12,16-18] and lack of consensus on how to conduct process evaluations of school-based interventions. Evaluations of school-based HPIs often assess one specific intervention within one school, or a specific single theme (e.g. a physical activity) that has been broadly implemented in many schools, which can make generalization to other theme areas or settings challenging.

In addition, obtaining objective data to measure the success of an intervention (i.e. expired carbon monoxide for tobacco control programs, pedometer data for physical activity interventions)[19] can be challenging and expensive. Perceptions of success, especially among decision makers within the school, may be equally if not more important measures of success, since perceived success may be a key driver in the decision to sustain an HPI within the school.[20]

Key features of successful HPIs identified to date include noncurricular approaches, playground interventions, after-school sessions and daily classroom refreshers.[12] In general, community-level interventions including those that are school-based should incorporate knowledge, beliefs and attitudes training while promoting healthy behaviour, since these features are related to intervention success regardless of theme area.[12,21] Further, emerging implementation science literature indicates that factors related to both the school (e.g. organizational context, leadership) and the intervention (e.g. partnerships, planning and implementation processes) are associated with implementation fidelity and effectiveness of HPIs.[22,23] Finally, school principals are key players in the school environment and instrumental in HPI implementation and sustainability.[24] Because they are knowledgeable about their school, the interventions offered and school staff opinions, their perception of HPI effectiveness is a key indicator of the potential usefulness of HPIs.

An increased understanding of what contributes to successful HPIs regardless of theme or setting could help school boards, school staff and the community increase autonomy in developing, selecting, implementing and evaluating interventions that align with school-specific needs. Further, identifying modifiable and nonmodifiable factors that contribute to or hinder the perceived success of HPIs could facilitate adaptations, improve implementation and contribute to the sustainability of school-based HPIs. Overall, offering schools evidence-informed interventions could increase the potential for HPI effectiveness and remove some of the guesswork from choosing an appropriate HPI.

Our objective in this study was to identify factors in three domains associated with perceived success of HPIs among elementary school principals. The three domains were school characteristics, characteristics of the HPI and factors related to planning and implementing HPIs. This project was undertaken as part of Project Prome*SS*, a cross-sectional survey of school principals and/or nominated staff members in a convenience sample of public elementary and high schools in Quebec.

Conceptual model

Delivery of school-based HPIs in Project Prome*SS* was envisioned based on a conceptual model guided by Rogers' Diffusion of Innovations Theory,[25,26] which explains how and why innovations (e.g. new HPIs) are adopted by schools. Specifically, the decision process is influenced by the characteristics of the innovation, the individuals involved and the organization implementing the innovations.[26] The process described by Rogers comprises four phases: planning, implementation, sustainability and scale-up. The PromeSS conceptual model[27] depicts these four phases in the context of school-based HPI delivery while illustrating environmental influences at the school, neighbourhood and societal levels. During planning, the school matches its needs with an existing intervention or develops an intervention *de novo*. During implementation, the intervention is delivered to students and may be modified. If an intervention is deemed unsuccessful, it can be terminated at any point. Interventions deemed successful may be renewed, become embedded in the school (i.e. sustained) or scaled up.[27] In this study, this model guided the selection of potential correlates of HPI perceived success.

Methods

Data were drawn from Project Prome*SS*. The sampling frame comprised all 1795 elementary and 436 high schools in 69 school boards across Quebec in 2016. Our analytical sample was restricted to elementary schools, since high schools differ markedly in student population, health issues perceived as important by school principals, and relevant HPI content and delivery methods. School board approval was obtained in 32 of 69 eligible school boards (46%), and 594 elementary schools (i.e. 33% of all elementary schools in Quebec) within the 32 school boards were eligible for recruitment. Private schools, schools serving only students with intellectual impairments or learning difficulties, and schools with fewer than 30 students were excluded because they are not assigned a school deprivation indicator. Contact was established with 291 of the 594 eligible elementary schools (49%); 171 of 291 eligible schools (59%) provided verbal assent and completed the interview.

Detailed data collection procedures are described elsewhere.[26] Briefly, schools were mailed or emailed a letter of introduction advising them of an upcoming telephone contact by the team. One week later, principals were contacted to confirm that they had worked in their current school longer than six months, and to solicit participation. If unavailable, the school principal nominated a vice-principal (n = 7/171) or another staff member (n = 5/171) to complete the interview.

Data were collected from 2016 to 2019 in two-part, structured telephone interviews

(median length 52.0 minutes) administered by trained interviewers in English or French. Participants provided data on school characteristics (i.e. school neighbourhood, funding from external sources, student demographics, perceived importance of specific student health issues), participant characteristics (i.e. sex, age, position, years working in the school) and availability of HPIs and extracurricular activities.

HPIs were defined as activities complementary to the educational curriculum offered to all students during class time at no cost, for which student attendance is mandatory. Information on HPIs for selected health theme areas is available elsewhere.[26] HPI availability was measured by asking: "In the past year, has your school offered any health-promoting interventions in which participation is expected at the group, class, grade, or school level to address …?", followed by a list of eight themes (physical activity/active living, sex education, healthy eating, bullying/exclusion, personal safety/injury prevention, mental health and well-being, oral health, tobacco control). Response options were Yes or No.

Participants were then asked to select one HPI offered within the last three years in order to respond to specific questions related to planning, implementing and sustaining that specific HPI. HPIs mandated by the government were ineligible for this section of the questionnaire. Questionnaire items were developed *de novo* or drawn or adapted from questionnaires used in previous work.[28] A retired school principal with more than 30 years' experience working in Quebec schools was centrally instrumental in developing the questionnaires. English and French questionnaires were pilot-tested by asking nine retired principals to narrate their thought processes as they interpreted the questions and formulated responses.

Ethics approval

Ethics approval was obtained from the Centre hospitalier de l'Université de Montréal (CHUM) Ethics Review Committee (2013-4130, CE 12.307).

Study variables

HPI theme area addressed by the HPI selected for in-depth questions was measured by asking: "What aspect(s) of your students' health and well-being does [name of intervention] primarily address?", followed by a list of 12 theme areas (smoking prevention, tobacco control education, aggressive behaviour, mental health [e.g. anxiety], bullying/cyberbullying, physical activity, healthy eating, addiction prevention, personal hygiene, puberty, personal safety/injury prevention, oral health). Descriptions of selected HPIs for each theme area are reported elsewhere.[26]

Perceived success of the HPI selected was measured using four items: (1) [intervention] met all objectives; (2) abandoning [intervention] had/would have a negative effect on the students; (3) [intervention] had a positive impact on students; and (4) animators enjoyed working on [intervention]. Participants responded to each item using a 5-point Likert-type response scale ranging from 1 (strongly disagree) to 5 (strongly agree). To create a mean score, responses were summed and divided by the number of items to which participants responded. Cronbach alpha for the score was 0.7. To provide evidence for convergent construct validity, we correlated perceived success against perceived permanence of the HPI (not at all, moderately, very permanent); the correlation coefficient was 0.27 ($p < 0.01$). Although perceived success does not measure whether the intervention actually resulted in behaviour change, it is a relevant indicator, since school principals who perceive an HPI as successful are more likely to invest resources and effort in its sustainability and in implementing other HPIs.[29]

Potential correlates of perceived success of the HPI were selected based on factors known to be associated with successful HPIs[30] and on availability of data in Prome*SS*. These included 11 school-related variables, namely number of students (range 37.0–889.0); number of (full- and part-time) teachers (range 5.0–58.0); language of instruction (English, French); percent of students in nutrition support program (range 0.0–100.0); school neighbourhood (urban, suburban, rural); teacher turnover (several, some, few, none); parent/community engagement in school (range 1.8–5.0); school/teacher commitment to student health (range 1.1–4.9); school physical environment (range 1.1–5.0); and principal leadership (range 2.6–5.0).

In addition, each school was ranked according to the 2016-2017 school deprivation indicator,[31] which is a composite score based on data for each student within the school reflecting whether the mother had completed high school and whether both parents were employed full-time. Scores ranged from 1 (lowest deprivation) to 10 (highest deprivation) and for descriptive purposes, schools were grouped into three categories: schools serving very advantaged students (i.e. school deprivation score = 1–3), those serving moderately advantaged students (4–7) and those serving disadvantaged (8–10) students.

Eight potential correlates related to the structural characteristics of the HPI were investigated: number of years HPI had been available in the school (range 1–43); number of competencies addressed by HPI (range 1–6); grades that received the HPI (yes/no for every grade); HPI was … a special event (yes/no), a pedagogical activity (yes/no) or a program (yes/no); number of learning strategies used in the HPI (range 0–4); and whether the school had a primary partner for the HPI (yes/no).

Nine potential correlates related to planning/implementing the HPI were studied: presence of implementation team leader (yes/no); number of implementation team members (range 2–42); HPI modified prior to implementation (yes/no); HPI modified during implementation (yes/no); school preparedness (range 1.0–5.0); program champion (at adoption or implementation stage; yes/no); number of types of evaluations conducted (range 0–7); school board involved in implementation (yes/no); and whether HPI was developed de novo (by the school) or adopted/adapted from an existing HPI.

Participant characteristics included sex, age, current position in school (principal, vice-principal, teacher), highest level of education completed, number of years of experience in current school (range 1.0–10.0) and number of years at current position (range 1.0–10.0).

Appendix A (available upon request) describes each potential correlate in detail, including questionnaire item(s), response options, coding for analysis and Cronbach alpha for scales.

Data analysis

We used descriptive statistics to characterize study schools and participants. Means

and standard deviation are presented for normally distributed variables, and medians and interquartile ranges for variables that were not normally distributed. To avoid issues of multiple testing, each potential correlate was investigated independently as a single hypothesis, and only two statistical tests (i.e. an unadjusted and a multivariable linear regression model) were performed for each potential correlate.[32,33] All multivariable models were adjusted for number of students, language of instruction, school neighbourhood and school deprivation. We did not test an omnibus model including all potential correlates, since this approach can be affected by an underdeveloped understanding of the possible relationships across all variables (especially in a cross-sectional study design), which can result in bias from over- or unnecessary adjustment. Data were analyzed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). All statistical tests were two-sided, with the significance level set at 0.05.

Results

School characteristics

The study sample included 163 elementary schools (i.e. for which data on perceived success for their chosen HPI were available) that were similar to all eligible elementary schools in Quebec (n = 1795). Specifically, 21% of schools in our sample served very advantaged students versus 24% of all eligible elementary schools; 44% versus 39% served moderately advantaged students and 36% versus 38% served disadvantaged students.[31] French was the official language in 83% of school boards in our sample versus 90% overall. The median number of students per school (n = 267) was similar to that in all eligible schools (n = 259).[26] One-quarter (25%) of study schools were located in urban neighbourhoods, 36% were suburban and 40% were rural. School principals reported French as the mother tongue of 98% of students. Finally, 42% of participants reported high teacher turnover and 22% reported high school principal turnover.

Participant characteristics

Sixty-nine percent of participants were female and almost all (97%) were the school principal. Mean (SD) age was 47.3 (7.4) years (range 30–60 years). Participants had worked a mean (SD) of 3.4 (2.6) years (range 1–10) in their current school and 7.1 (3.4) years in their current position (range 1–10).[26]

Description of HPIs

HPIs selected for in-depth questions by the participant often addressed more than one theme area (e.g. physical activity and healthy eating; bullying and mental health). Among the 171 schools studied, 154 different HPIs were reported. More than half (58%) of HPIs addressed physical activity (e.g. daily 15-minute walk for students and staff); 43% addressed healthy eating (e.g. healthy cooking workshop animated by the teacher); 30% addressed personal safety and/or injury prevention (e.g. workshop in conjunction with improved policy to promote safe walking and biking to school); 26% focussed on bullying; 25% targeted aggressive behaviour (e.g. in-class conversations animated by teachers or psychosocial staff); and 21% addressed mental health (e.g. teacher-led workshop to teach young children to verbalize emotions through storytelling). Few HPIs selected by participants for in-depth questions addressed personal hygiene (9%), puberty (6%), addiction prevention (5%), oral health (3%) or tobacco prevention and education (2%).

Perceived success of HPIs

Scores for perceived success of HPIs ranged from 2.3 to 5.0, with a mean (SD) of 4.3 (0.5). The assumption of normality in the distribution of scores was supported.[34]

Correlates of perceived success

Results of the multivariable linear regression analyses adjusting for number of students, language of instruction, school neighbourhood and school deprivation indicated that four of 11 school characteristics were associated with perceived success of the HPI including lower teacher turnover, school physical environment, school/teacher commitment to student health and principal leadership (Table 1). No variable describing structural characteristics of the HPI was associated with perceived success (Table 2). Finally, only one of nine variables related to HPI planning or implementation was related to perceived success. Specifically, being a developer (vs. an adopter) of the HPI related to higher scores of perceived success (Table 3).

Discussion

In this study of Quebec elementary schools, we drew on our conceptual model depicting key elements to consider in the delivery of school-based HPIs, to select potential correlates of perceived success. Although school principals generally perceived HPIs as highly successful, there was variability in perceived success scores, and five factors emerged as correlates. These pertained to school characteristics and to planning and implementing HPIs, but none of the HPI structural characteristics investigated were retained.

School characteristics

Because both the environment and the "actors" involved in a school-based intervention can influence how an intervention is delivered and whether it produces the intended effects,[35] we investigated the context of health promotion programming[36] according to school-level correlates. Among 11 variables describing school characteristics, four (i.e. lower teacher turnover, school physical environment, school/teacher commitment to student health, principal leadership) were associated with perceived HPI success.

First, frequent turnover of school staff could challenge HPI implementation because of lack of continuity, changes in staff priorities and motivation and loss of the "corporate history."[37] It may be prudent for educators and HPI developers to incorporate training larger numbers of staff in HPI implementation, and to foster institutionalizing HPIs into the school curriculum.[24]

Second, as in earlier studies,[24] HPIs were perceived as more successful when components of school culture, including school physical environment, school/teacher commitment to student health and principal leadership, were rated higher. School culture represents the shared beliefs and norms of the school[38] and encompasses the operational processes and motivations that guide HPI delivery. Availability of equipment and space can provide school staff with greater latitude in their HPI choice, increasing their probability of selecting an intervention that fits with the school context.

TABLE 1
Unstandardized beta (β) coefficients and 95% CIs from linear regression models for the association between school characteristics and perceived success of school-based health-promoting interventions (HPIs), Project PromeSS, 2016–2019 (n = 163)

	n	Perceived success Mean (SD)	β_{crude} (95% CI)	$\beta_{adjusted}$[a] (95% CI)
No. students[b]			0.00 (−0.04, 0.04)[c]	0.00 (−0.06, 0.06)[c]
37–267	81	4.25 (0.51)		
268–889	81	4.26 (0.52)		
No. teachers[b]			0.00 (−0.01, 0.01)	−0.01 (−0.03, 0.02)
5–18	67	4.26 (0.49)		
19–37	81	4.24 (0.51)		
≥ 38	15	4.31 (0.66)		
Language of instruction				
French	136	4.22 (0.51)	ref	ref
English	27	4.43 (0.49)	0.21 (0.00, 0.42)	0.21 (0.00, 0.43)
% of students in nutrition support program[b]			0.00 (0.00, 0.00)	0.00 (−0.01, 0.00)
0	105	4.29 (0.53)		
1–100	50	4.18 (0.49)		
School neighbourhood				
Urban/suburban	98	4.25 (0.51)	ref	ref
Rural	65	4.25 (0.52)	0.00 (−0.08, 0.08)	0.01 (−0.10, 0.11)
School deprivation[b]			0.00 (−0.03, 0.03)	0.00 (−0.04, 0.03)
High	34	4.32 (0.47)		
Moderate	71	4.21 (0.58)		
Low	58	4.27 (0.46)		
Teacher turnover[b]			**0.12 (0.03, 0.21)**	**0.13 (0.04, 0.21)**
Several	17	4.03 (0.54)		
Some	52	4.21 (0.52)		
Few	61	4.22 (0.49)		
None	31	4.46 (0.46)		
Parent/community engagement in school[b]			0.02 (−0.10, 0.15)	0.04 (−0.09, 0.18)
< 3.8	70	4.24 (0.49)		
≥ 3.8	93	4.26 (0.54)		
School physical environment[b]			**0.21 (0.08, 0.34)**	**0.20 (0.07, 0.33)**
< 3.6	88	4.15 (0.50)		
≥ 3.6	75	4.37 (0.51)		
School/teacher commitment to student health[b]			**0.25 (0.11, 0.39)**	**0.28 (0.13, 0.42)**
< 4.0	46	4.08 (0.55)		
≥ 4.0	117	4.32 (0.48)		
Principal leadership[b]			**0.19 (0.03, 0.36)**	**0.20 (0.04, 0.37)**
< 3.9	86	4.14 (0.50)		
≥ 3.9	60	4.36 (0.48)		

Abbreviations: CI, confidence interval; SD, standard deviation.

Note: Bold type indicates confidence intervals that do not include the null. Totals do not always sum to 163 because of missing data. The beta coefficient represents the change in perceived success for every 1-unit change in the correlate.

[a] All models adjusted for number of students, language of instruction, school neighbourhood and school deprivation.

[b] Responses for continuous potential correlates were categorized for descriptive purposes, and the mean (SD) was computed for each group. However, these variables were retained as continuous in the modelling.

[c] The estimate represents a change in the number of students per 100.

Third, school/teacher commitment to student health, which reflects emphasis on and commitment to health promotion by school staff, may positively influence how HPIs are perceived within schools where staff believe in their relevance.[39]

Finally, because school principals are central in guiding staff towards objectives, obtaining resources, distributing responsibilities and solving conflicts,[40] their leadership can be key. Multiple studies stress the need for strong leadership to facilitate HPI delivery.[41-45]

HPI characteristics

Roger's diffusion theory[25] posits that perceptions of the relative advantage, compatibility, complexity, trialability and observability of an intervention are key in selecting and evaluating interventions. We investigated characteristics of HPIs in two categories—structural characteristics, and planning and implementation.

Structural characteristics represent features of the HPI such as target audience and learning strategies used to transmit health knowledge and effect behaviour

TABLE 2
Unstandardized beta (β) coefficients and 95% CIs from linear regression models for the association between eight structural characteristics of school-based health-promoting interventions (HPIs) and perceived success of HPI, Project PromeSS, 2016–2019 (n = 163)

	n	Perceived success Mean (SD)	β_{crude} (95% CI)	$\beta_{adjusted}$ [a] (95% CI)
No. years HPI in school[b]			0.08 (−0.04, 0.20)	0.10 (−0.03, 0.23)
1	32	4.17 (0.56)		
2–5	85	4.25 (0.53)		
≥ 6	36	4.33 (0.43)		
No. of competencies addressed in HPI[b]			0.05 (0.00, 0.10)	0.04 (−0.02, 0.09)
1	58	4.28 (0.50)		
2	42	4.09 (0.56)		
3–6	63	4.34 (0.48)		
All grades received HPI				
No	56	4.36 (0.46)	ref	ref
Yes	107	4.20 (0.53)	−0.16 (−0.33, 0.01)	−0.15 (−0.32, 0.02)
HPI was a special event[c]				
No	114	4.29 (0.54)	ref	ref
Yes	49	4.17 (0.45)	−0.13 (−0.30, 0.05)	−0.11 (−0.29, 0.07)
HPI was a pedagogical activity[c]				
No	114	4.25 (0.55)	ref	ref
Yes	49	4.25 (0.42)	0.00 (−0.18, 0.17)	0.01 (−0.17, 0.19)
HPI was a program[c]				
No	93	4.22 (0.46)	ref	ref
Yes	70	4.29 (0.58)	0.07 (−0.09, 0.23)	0.04 (−0.13, 0.20)
No. learning strategies[b,d]			0.05 (−0.03, 0.14)	0.04 (−0.05, 0.13)
1	71	4.21 (0.57)		
2	55	4.23 (0.48)		
3	26	4.41 (0.41)		
4	11	4.26 (0.52)		
School worked with a partner				
No	50	4.31 (0.56)	ref	ref
Yes	113	4.23 (0.49)	−0.09 (−0.26, 0.09)	−0.09 (−0.26, 0.09)

Abbreviations: CI, confidence interval; SD, standard deviation.

Note: Bold type indicates confidence intervals that do not include the null. Totals do not always sum to 163 because of missing data. The beta coefficient represents the change in perceived success for every 1-unit change in the correlate.

[a] All models adjusted for number of students, language of instruction, school neighbourhood and school deprivation.

[b] Responses for continuous potential correlates were categorized for descriptive purposes, and the mean (SD) was computed for each group. However, these variables were retained as continuous in the modelling.

[c] Participants were instructed to choose all responses that applied to the questionnaire item: [Name of intervention] was a … (1) special event (e.g. health fair, guest speaker at an assembly, etc.) (specify); (2) pedagogical activity; (3) learning and evaluation situation; (4) program (specify); (5) other (specify).

[d] Participants were instructed to choose all responses that applied to the questionnaire item: What type of learning strategy was used for [name of intervention]? (1) lecture strategies: presentations, demonstrations; (2) individual work: independent practice; (3) interactive teaching strategies: group discussion, role-play, modelling; (4) social constructivist teaching strategies: peer education, tutoring, collaborative and cooperative learning; (5) other (specify).

change. Among eight variables in this category, none were associated with perceived success, although other studies do report that these features are associated with HPI effectiveness. In a systematic review, school-based substance use programs were more effective when focussed on competencies including social skills, self-control and problem-solving.[46] A review of obesity prevention interventions for preschool children identified interactive learning strategies, such as modelling, as key.[47] In our study, rather than focus on HPIs targeting a specific theme, we assessed a broad range of correlates of perceived success of diverse HPIs. Regardless of this heterogeneity, principals regarded most interventions as highly successful, suggestive that correlates other than structural factors might contribute more to perceived success.

We investigated nine characteristics related to planning and implementing HPIs. Based on Rogers' diffusion theory,[25] planning is the first phase of HPI delivery, comprising identification of a need for the HPI in the school and learning about alternate HPIs that can respond to that need.[48] Schools may seek information on existing interventions, be solicited by HPI developers or develop an HPI themselves. Implementation comprises delivering the intervention to students and may involve continuous adjustment to the school context.[27] In this study, the only planning and implementation characteristic associated with perceived success was that the "school developed its own HPI." Staff may feel more ownership of HPIs developed in-house, which may lead to higher levels of commitment and trust in expected benefits.[35,49] It is possible that in-house development produced HPIs better tailored to the school context, since school personnel likely have a well-developed understanding of their students' needs.

Strengths and limitations

Strengths of this study include that it examines numerous correlates of perceived

TABLE 3
Unstandardized beta (β) coefficients and 95% CIs from linear regression models for the association between nine factors related to planning/implementing school-based health-promoting interventions (HPIs) and perceived success of HPI, Project PromeSS, 2016–2019 (n = 163)

	n	Perceived success Mean (SD)	β_{crude} (95% CI)	$\beta_{adjusted}$[a] (95% CI)
No. implementation team members[b,c]			−0.01 (−0.03, 0.01)	−0.01 (−0.03, 0.01)
2–4	50	4.31 (0.39)		
5–42	41	4.18 (0.54)		
Implementation team leader[c]				
No	27	4.17 (0.45)	ref	ref
Yes	65	4.28 (0.48)	0.12 (−0.10, 0.33)	0.10 (−0.12, 0.31)
HPI modified prior to implementation				
No	96	4.23 (0.49)	ref	ref
Yes	48	4.30 (0.52)	0.07 (−0.11, 0.24)	0.05 (−0.13, 0.23)
HPI modified during implementation				
No	65	4.23 (0.51)	ref	ref
Yes	98	4.27 (0.52)	0.04 (−0.13, 0.20)	0.04 (−0.12, 0.21)
School preparedness[b]			**0.09 (0.01, 0.17)**	0.08 (0.00, 0.16)
0–2.75	90	4.17 (0.55)		
≥ 2.76	73	4.36 (0.45)		
Program champion				
No	55	4.14 (0.54)	ref	ref
Yes	106	4.30 (0.49)	0.16 (−0.01, 0.33)	0.14 (−0.03, 0.32)
No. of types of evaluation[b,d]			0.05 (0.00, 0.10)	0.05 (0.00, 0.11)
0	3	4.06 (0.82)		
1	14	4.27 (0.69)		
2	21	4.17 (0.48)		
3	37	4.12 (0.55)		
4	44	4.30 (0.47)		
5	21	4.23 (0.45)		
6–7	23	4.48 (0.43)		
School board involved in implementation				
No	116	4.20 (0.52)	ref	ref
Yes	30	4.36 (0.40)	0.16 (−0.04, 0.37)	0.18 (−0.03, 0.38)
School				
Developed HPI	97	4.35 (0.45)	ref	ref
Adopted HPI	66	4.12 (0.57)	**−0.23 (−0.39, −0.07)**	**−0.24 (−0.40, −0.08)**

Abbreviations: CI, confidence interval; SD, standard deviation.

Note: Bold type indicates confidence intervals that do not include the null. Totals do not always sum to 163 because of missing data. The beta coefficient represents the change in perceived success for every 1-unit change in the correlate.

[a] All models adjusted for number of students, language of instruction, school neighbourhood and school deprivation.

[b] Responses for continuous potential correlates were categorized for descriptive purposes, and the mean (SD) was computed for each group. However, these variables were retained as continuous in the modelling.

[c] Only HPIs with team members responded to this question.

[d] Participants were instructed to choose all responses that applied for the questionnaire item: Did your school do any of the following to evaluate [name of intervention]? (1) Hold regular meetings; (2) Obtain feedback from the [name of intervention] animators; (3) Document the extent to which implementation was carried out in accordance with the plan; (4) Document the number of students participating in the [name of intervention]; (5) Document the barriers and facilitators to implementation; (6) Formally evaluate the outcomes of the [name of intervention]; (7) Other (specify).

success across a wide variety of HPIs. In addition, although Prome*SS* included a convenience sample of schools, which could limit generalizability,[50] the characteristics of Prome*SS* schools resembled those of all eligible elementary schools in Quebec.

Limitations include that, although responses from a single person within a school may not provide an accurate portrayal of the organizational perspective, data collection from multiple respondents within the same school was not feasible. However, the Prome*SS* questionnaire was sent to participants in advance of the interview so that participants could consult their staff in preparation for the interview. Our measure of perceived success was created *de novo*. Until its validity and reliability are established, the interpretation of absolute differences between scores remains uncertain. Further, responses were right-skewed (i.e. more participants perceived success favourably), which limited variability and may have rendered detection of correlates more difficult. Recall error could have resulted in misclassification bias in the observed associations. Finally, the precision of estimates in Prome*SS* was limited because of the relatively small sample size.

Conclusion

School personnel in elementary schools generally perceived that school-based HPIs

are highly successful. Correlates of perceived success include low teacher turnover, positive school physical environment, school/teacher commitment to student health, principal leadership and developing the HPI *de novo*. If replicated in other independent studies, these factors should be considered by HPI developers and school personnel when planning and implementing HPIs in schools.

Acknowledgements

JOL held a Canada Research Chair in the Early Determinants of Adult Chronic Disease from 2006 to 2021. JK and TR both hold a doctoral scholarship from the Quebec Order of Nurses (Ordre des infirmières et infirmiers du Québec). EOL holds a post-doctoral salary award from the Fonds de recherche du Québec—Santé (FRQ-S). Funding sources were not involved in the study design, data collection, analysis or interpretation, or the preparation of the manuscript for publication.

Conflicts of interest

The authors have no competing interests.

Funding

This work was supported by Quebec Ministry of Health and Social Services.

Authors' contributions and statement

All authors contributed to conceptualization of the study objectives and analytic plan, interpreted the results and reviewed and revised the manuscript. EOL conducted the analyses, reviewed the literature and drafted the manuscript. JK developed the school culture variables, wrote the discussion and reviewed the manuscript. AP conducted analyses and reviewed the manuscript. TR contributed to analysis of the results and reviewed the manuscript. JOL developed and oversaw all aspects of Project Prome*SS* including its conceptualization, funding, design and data collection. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

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

References

1. Wylie A, Holt T. Health promotion in medical education: from rhetoric to action. Oxford (UK): Radcliffe Publishing; 2010. 314 p.

2. World Health Organization, United Nations Educational, Scientific and Cultural Organization. Making every school a health-promoting school: global standards and indicators for health-promoting schools and systems. Geneva (CH): World Health Organization; 2021. 50p. Available from: https://www.who.int/publications/i/item/9789240025059

3. Howie EK, McVeigh JA, Smith AJ, et al. Physical activity trajectories from childhood to late adolescence and their implications for health in young adulthood. Prev Med. 2020;139:106224. https://doi.org/10.1016/j.ypmed.2020.106224

4. Riglea T, Doré I, O'Loughlin J, Belanger M, Sylvestre MP. Contemporaneous trajectories of physical activity and screen time in adolescents. Physiol Nutr Metab. 2021;46(6):676-84. https://doi.org/10.1139/apnm-2020-0631

5. Tucker JS, Ellickson PL, Orlando M, Martino SC, Klein DJ. Substance use trajectories from early adolescence to emerging adulthood: a comparison of smoking, binge drinking, and marijuana use. J Drug Issues. 2005;35(2):307-32. https://doi.org/10.1177/002204260503500205

6. Langford R, Bonell CP, Jones HE, et al. The WHO health promoting school framework for improving the health and well-being of students and their academic achievement. Cochrane Database Syst Rev. 2014(4):CD008958. https://doi.org/10.1002/14651858.CD008958.pub2

7. Naylor P-J, McKay HA. Prevention in the first place: schools a setting for action on physical inactivity. Br J Sports Med. 2009;43(1):10-13. https://doi.org/10.1136/bjsm.2008.053447

8. Ministère de l'Éducation du Loisir et du Sport. Quebecers on the move! Policy on physical activity, sport and recreation. Québec (QC): Gouvernement du Québec; 2017. 44 p. Available from: http://www.education.gouv.qc.ca/fileadmin/site_web/documents/loisir-sport/Policy_Quebecers_on_the_move.pdf

9. Ministère de la Santé et des Services sociaux. Programme national de santé publique 2015–2025. Québec (QC): Gouvernement du Québec; 2015. 88 p. Available from: https://publications.msss.gouv.qc.ca/msss/fichiers/2015/15-216-01W.pdf

10. Ministère de l'Éducation et de l'Enseignement supérieur. Learning content in sexuality education. Québec (QC): Gouvernement de Québec; 2018. 24 p. Available from: http://collections.banq.qc.ca/ark:/52327/3547856.

11. National Assembly of Québec. *An act to prevent and stop bullying and violence in schools*. (SQ 2012, c 19). Québec (QC): Gouvernement du Québec; 2012. Available from: http://www2.publicationsduquebec.gouv.qc.ca/dynamicSearch/telecharge.php?type=5&file=2012C19A.PDF

12. Jepson R, Harris F, MacGillivray S, Kearney N, Rowa-Dewar NJ. A review of the effectiveness of interventions, approaches and models at individual, community and population level that are aimed at changing health outcomes through changing knowledge attitudes and behaviour. London (UK): National Institute for Health and Clinical Excellence; 2006. 218 p. Available from: https://www.nice.org.uk/guidance/ph6/evidence/behaviour-change-review-1-effectiveness-review-pdf-369664525

13. Fung C, Kuhle S, Lu C, et al. From "best practice" to "next practice": the effectiveness of school-based health promotion in improving healthy eating and physical activity and preventing childhood obesity. Int J Behav Nutr Phys Act. 2012;9:27. https://doi.org/10.1186/1479-5868-9-27

14. Langford R, Bonell C, Jones H, et al. The World Health Organization's Health Promoting Schools framework: a Cochrane systematic review and meta-analysis. BMC Public Health. 2015;15:130. https://doi.org/10.1186/s12889-015-1360-y

15. Chilton R, Pearson M, Anderson R. Health promotion in schools: a scoping review of systematic reviews. Health Educ. 2015;115(3/4):357-76. https://doi.org/10.1108/HE-03-2014-0033

16. Mũkoma W, Flisher AJ. Evaluations of health promoting schools: a review of nine studies. Health Promot Int. 2004;19(3):357-68. https://doi.org/10.1093/heapro/dah309

17. Speller V, Learmonth A, Harrison D. The search for evidence of effective health promotion. BMJ. 1997;315(7104):361-3. https://doi.org/10.1136/bmj.315.7104.361

18. World Health Organization. Health promotion evaluation: recommendations to policy-makers. Report of the WHO European Working Group On Health Promotion Evaluation. Copenhagen (Denmark): WHO Regional Office for Europe; 1998. [Contract No.: EUR/ICP/IVST 05 01 03.] 12 p. Available from: https://apps.who.int/iris/handle/10665/108116

19. Bastian KA, Maximova K, McGavock J, Veugeleres P. Does school-based health promotion affect physical activity on weekends? And, does it reach those students most in need of health promotion? PLoS ONE. 2015;10(10):e0137987. https://doi.org/10.1371/journal.pone.0137987

20. Leger LS, Kolbe L, Lee A, McCall DS, Young IM. School health promotion. In: McQueen DV, Jones CM, editors. Global perspectives on health promotion effectiveness. New York (NY): Springer New York; 2007. p. 107-124.

21. Trickett EJ, Beehler S, Deutsch C, et al. Advancing the science of community-level interventions. Am J Public Health. 2011;101(8):1410-19. https://doi.org/10.2105/ajph.2010.300113

22. Lederer AM, King MH, Sovinski D, Seo D-C, Kim N. The relationship between school-level characteristics and implementation fidelity of a coordinated school health childhood obesity prevention intervention. J Sch Health. 2015;85(1):8-16. https://doi.org/10.1111/josh.12221

23. Rowling L, Samdal O. Filling the black box of implementation for health-promoting schools. Health Educ. 2011;111(5):347-66. https://doi.org/10.1108/09654281111161202

24. Forman SG, Olin SS, Hoagwood KE, et al. Evidence-based interventions in schools: developers' views of implementation barriers and facilitators. School Ment Health. 2009;1:26-36. https://doi.org/10.1007/s12310-008-9002-5

25. Rogers EM. Diffusion of innovations. New York (NY): The Free Press; 1983. 453 p.

26. Riglea T, Kalubi J, Sylvestre MP, et al. Social inequalities in availability of health-promoting interventions in Québec elementary schools. Health Promot Int. 2022;37(1). https://doi.org/10.1093/heapro/daab023

27. Mihalic SF, Irwin K. Blueprints for violence prevention: from research to real-world settings—factors influencing the successful replication of model programs. Youth Violence Juv Justice. 2003;1(4):307-29. https://doi.org/10.1177/1541204003255841

28. O'Loughlin J, Renaud L, Richard L, Gomez LS, Paradis G. Correlates of the sustainability of community-based heart health promotion interventions. Prev Med. 1998;27(5):702-12. https://doi.org/10.1006/pmed.1998.0348

29. Guldager JD, Andersen PT, von Seelen J, Leppin A. Physical activity school intervention: context matters. Health Educ Res. 2018;33(3):232-42. https://doi.org/10.1093/her/cyy012

30. Deschesnes M, Martin C, Hill AJ. Comprehensive approaches to school health promotion: how to achieve broader implementation? Health Promot Int. 2003;18(4):387-96. https://doi.org/10.1093/heapro/dag410

31. Ministère de l'Éducation et de l'Enseignement supérieur. Indices de défavorisation des écoles publiques, 2016-2017. Québec (QC): Ministère de l'Éducation et de l'Enseignement supérieur; 2017. 80 p. Available from: http://www.education.gouv.qc.ca/fileadmin/site_web/documents/PSG/statistiques_info_decisionnelle/Indices_PUBLICATION_20162017_final.pdf

32. Hagströmer M, Oja P, Sjöström M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. Public Health Nutr. 2006;9(6):755-62. https://doi.org/10.1079/phn2005898

33. Bender R, Lange S. Adjusting for multiple testing--when and how? J Clin Epidemiol. 2001;54(4):343-9. https://doi.org/10.1016/s0895-4356(00)00314-0

34. Osbourne JW, Waters EM. Four assumptions of multiple regression that researchers should always test. Pract Assess Res Eval. 2002;8:2. https://doi.org/10.7275/r222-hv23

35. May CR, Johnson M, Finch T. Implementation, context and complexity. Implement Sci. 2016;11:141. https://doi.org/10.1186/s13012-016-0506-3

36. Poland B, Frohlich KL, Cargo M. Context as a fundamental dimension of health promotion program evaluation. In: Potvin L, McQueen DV, Hall M, Salazar L, Anderson LM, Hartz ZM, editors. Health promotion evaluation practices in the Americas: values and research. New York: Springer; 2008. p. 299-317.

37. Darlington EJ, Simar C, Jourdan D. Implementation of a health promotion programme: a 10 year retrospective study. Health Educ. 2017;117(3):252-79. https://doi.org/10.1108/HE-09-2016-0038

38. Domitrovich CE, Bradshaw CP, Poduska JM, et al. Maximizing the implementation quality of evidence-based preventive interventions in schools: a conceptual framework. Adv Sch Ment Health Promot. 2008;1(3):6-28. https://doi.org/10.1080/1754730x.2008.9715730

39. Todd C, Christian D, Davies H, et al. Headteachers' prior beliefs on child health and their engagement in school based health interventions: a qualitative study. BMC Res Notes. 2015;8:161. https://doi.org/10.1186/s13104-015-1091-2

40. Paredes Campusano MJ. Le nouveau rôle du directeur d'établissement dans la formation continue des enseignants dans le cadre de la réforme du curriculum au primaire. Montréal (QC): Université du Québec à Montréal; 2006. 86 p.

41. Deschesnes M, Drouin N, Couturier Y. Schools' absorptive capacity to innovate in health promotion. J Health Organ Manag. 2013;27(1):24-41. https://doi.org/10.1108/14777261311311780

42. Gaudet J, Breton A. Recension des programmes d'intervention en milieu scolaire visant le mieux vivre ensemble : qu'est-ce qui favorise leur efficacité et leur implantation? Montréal (QC): Institut Pacifique; 2009. 84 p. Available from: http://programme-allie.ca/pdf/doc_recens_progr-2009.pdf

43. Lyon AR, Cook CR, Brown EC, et al. Assessing organizational implementation context in the education sector: confirmatory factor analysis of measures of implementation leadership, climate, and citizenship. Implement Sci. 2018;13(1):5. https://doi.org/doi:10.1186/s13012-017-0705-6

44. Moore GF, Littlecott HJ, Fletcher A, Hewitt G, Murphy S. Variations in schools' commitment to health and implementation of health improvement activities: a cross-sectional study of secondary schools in Wales. BMC Public Health. 2016;16:138. https://doi.org/10.1186/s12889-016-2763-0

45. Simard C, Deschesnes M. Recension des écrits publiés entre 2000 et 2009 sur les résultats d'évaluation des approches globales en promotion de la santé en contexte scolaire. Montréal (QC): Institut national de santé publique du Québec; 2011. 79 p. Available from: http://www.santecom.qc.ca/bibliothequevirtuelle/hyperion/9782550612087.pdf

46. Onrust SA, Otten R, Lammers J, Smit F. School-based programmes to reduce and prevent substance use in different age groups: what works for whom? Systematic review and meta-regression analysis. Clin Psychol Rev. 2016;44:45-59. https://doi.org/10.1016/j.cpr.2015.11.002

47. Summerbell CD, Moore HJ, Vögele C, et al. Evidence-based recommendations for the development of obesity prevention programs targeted at preschool children. Obes Rev. 2012;13(s1):129-32. https://doi.org/10.1111/j.1467-789X.2011.00940.x

48. Wisdom JP, Chor KHB, Hoagwood KE, Horwitz SM. Innovation adoption: a review of theories and constructs. Adm Policy Ment Health. 2014;41(4):480-502. https://doi.org/10.1007/s10488-013-0486-4

49. Gugglberger L, Inchley J. Phases of health promotion implementation into the Scottish school system. Health Promot Int. 2014;29(2):256-66. https://doi.org/10.1093/heapro/das061

50. Jager J, Putnick DL, Bornstein MH. II. More than just convenient: the scientific merits of homogeneous convenience samples. Monogr Soc Res Child Dev. 2017;82(2):13-30. https://doi.org/10.1111/mono.12296

https://doi.org/10.24095/hpcdp.42.9.04

Original quantitative research

Do school characteristics, based on the Comprehensive School Health framework, contribute to youth meeting national physical activity recommendations over time?

Melissa Pirrie, PhD (1); Valerie Carson, PhD (2); Joel A. Dubin, PhD (1,3); Scott T. Leatherdale, PhD (1)

This article has been peer reviewed.

Abstract

Introduction: Canadian youth are insufficiently active, and schools may play a role in promoting student physical activity (PA). Based on the Comprehensive School Health (CSH) framework, this study examined whether school characteristics are associated with secondary school students meeting national PA recommendations over time.

Methods: We used COMPASS survey data from 78 schools in Ontario and Alberta and 9870 Grade 9 and 10 students attending those schools. Students who provided two years of linked PA data (2013/14 and 2015/16) and gender were included. Multilevel analysis was conducted by gender, evaluating the relationship of school-level characteristics (guided by CSH) with students achieving all three PA recommendations after two years (≥ 60 min/day of moderate-to-vigorous PA, vigorous PA ≥ 3 days/week, strengthening activities ≥ 3 days/week).

Results: More than half (56.9%) of students achieving the PA recommendations at baseline were no longer achieving them after two years, and just a quarter (25.6%) of students not achieving the recommendations at baseline achieved them after two years. School-level factors were significantly associated with students achieving the recommendations, but these factors differed by student strata (i.e. by gender and baseline PA status). Generally, student access to equipment, public health partnerships and staff time for health were associated with increased odds of achieving the PA recommendations for certain students.

Conclusion: Modifications to school characteristics within CSH may play a role in supporting students in achieving or continuing to achieve the PA recommendations after two years. Further research is needed to better understand the underlying dynamics of the observed relationships.

Keywords: *youth, schools, multilevel regression modelling, gender, guidelines*

Highlights

- Approximately half of Grade 9 and 10 students were not meeting the national physical activity recommendations at baseline.
- There is evidence that inactive male and female students can transition to being active two years later, although this was observed in only a quarter of inactive students.
- School characteristics within the Comprehensive School Health framework may support both male and female students in becoming active or staying active.
- School-level factors associated with students achieving the physical activity recommendations were: partnering with public health, providing access to equipment during non-instructional time and receiving staff time for student health from the school board.

Introduction

The *Canadian 24-hour Movement Guidelines for Children and Youth,* released in 2016, recommend that youth accumulate an average of 60 minutes per day of moderate-to-vigorous physical activity (MVPA), participate in vigorous physical activity (VPA) at least three days per week and participate in muscle-and-bone-strengthening activities at least three days per week.[1,2] In 2017, objective data from the Canadian Health Measures Survey[3] suggested that only 31% of Canadian youth were achieving the recommended average of 60 minutes per day of MVPA. The proportion of youth not achieving national physical activity (PA) recommendations is a major public health concern, as PA is a modifiable health behaviour that has many immediate health benefits (e.g. improved cardiovascular health, decreased anxiety) and is associated with long-term prevention of chronic disease and some cancers.[4-6]

For secondary school students, an additional concern is that those meeting the PA recommendations in Grades 9 and 10 may not continue achieving them a few years later. Evidence consistently shows that PA declines throughout adolescence,

Author references:

1. School of Public Health Sciences, University of Waterloo, Waterloo, Ontario, Canada
2. Faculty of Kinesiology, Sport, and Recreation, University of Alberta, Edmonton, Alberta, Canada
3. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

Correspondence: Melissa Pirrie, School of Public Health Sciences, University of Waterloo, 200 University Ave. West, Waterloo, ON N2L 3G1; Tel: 519-580-6489; Email: melissa.pirrie@uwaterloo.ca

especially in girls,[3,7] with lower PA observed in Canadian youth (aged 12–17 years) compared to children (aged 6–11 years).[3] Since PA during youth significantly predicts PA behaviours during adulthood,[6,8] preventing a drop-off during this critical developmental period is important. Similarly, providing an environment that encourages students who are not already meeting the PA recommendations in Grades 9 and 10 to improve their PA behaviours may help put them on a healthier behavioural trajectory before entering early adulthood.

Ecological theory suggests that the individual's context (e.g. family, school, neighbourhood) can impact health behaviours.[9] As schools are an integral part of the students' PA environment,[10-12] this is a context frequently targeted to improve youth PA behaviour.[13] However, many investigator-initiated PA interventions in schools struggle to be sustained post-research due to a lack of capacity and resources.[13] Taking a different approach, the evaluation of natural experiments through longitudinal data systems can be used to identify which programs, policies and resources may have been successful in achieving higher PA in some schools to inform what may be feasible in other schools.[14]

The Comprehensive School Health (CSH) framework provides guidance on four inter-related components that schools can target to improve health behaviours, such as PA.[15,16] The CSH components are: (1) social and physical environment; (2) teaching and learning; (3) partnerships and services; and (4) policies.[15,16] Many cross-sectional studies have examined the associations between school characteristics found within CSH and PA behaviours, finding students are more physically active if they attend schools with higher social support,[17,18] established partnerships,[19] extracurricular PA opportunities,[20] PA facilities available,[21] student access to facilities and equipment outside of instructional time[21] and funding for PA-related resources.[22] However, there has been no longitudinal study examining how multiple school characteristics representing each component of the CSH framework are associated with students achieving the national PA recommendations over time.

The objective of this exploratory study was to fill an existing evidence gap by using longitudinal data to determine (1) the proportion of students achieving all three PA recommendations at baseline who were still achieving them two years later; (2) the proportion of students not achieving all three PA recommendations at baseline who were achieving them two years later; and (3) the school characteristics associated with achieving all three recommendations two years later for these two groups of students. Since gender has been identified as a modifier of the association between school characters and student PA,[23] these relationships were evaluated for each gender separately.

Methods

Ethics approval

Ethics approval was received by the University of Waterloo Office of Research Ethics (Project: 30188), the University of Alberta Research Ethics Board (Project: 00040729) and each school board and/or school as required.

Design

COMPASS is a prospective cohort study (2012–2021) following a large sample of Canadian students in Grades 9 to 12 and the schools they attend. COMPASS was designed as a platform to conduct natural experiments to evaluate the impact of school-level prevention program and policy changes on student health behaviours.[24]

The current study examined data from 78 schools in Ontario (n = 69) and Alberta (n = 9) that provided data in both Year 2 (Y_2; 2013/14) and Year 4 (Y_4; 2015/16); this is the period in which COMPASS had the largest initial intake of schools to its cohort and the schools completed their baseline COMPASS School Policies and Practices Questionnaire (2012/13 or 2013/14). Of the 78 schools, 72 were public and 6 were private schools with a traditional education setting; all private schools were located in Ontario.[25-27] Although COMPASS is a rolling cohort that includes all grades of secondary school students, the current study only included students in Grade 9 or 10 as of Y_2 since students in Grades 11 and 12 would have graduated and left the cohort by Y_4. The current study included 9870 students (9239 in Ontario and 631 in Alberta) who provided gender and PA data in both Y_2 and Y_4, and provided their grade for at least one of the years. There were no significant differences in students included or excluded due to missing outcome data according to chi-square tests (data available upon request).

Recruitment

Detailed school recruitment methods are available.[24-27] In short, school boards that allow active-information, passive-consent protocols were approached to participate in COMPASS. With school board approval, individual schools were recruited and information was sent to students and guardians, with multiple mechanisms to opt out. Students could also opt to not complete the survey on data collection day and could withdraw their survey at any time. Seventy-nine percent of eligible students from these 78 schools participated in Y_2 and 80% of eligible students participated in Y_4. Missing students were primarily due to absenteeism or classroom spares on the day of data collection; less than 1% of students opted out (themselves or their parent or guardian).

Data collection procedures

On each school's data collection day, a COMPASS staff member was onsite to ensure study fidelity and identify the most appropriate individual to complete the school-level survey. Students completed the survey during a class period and sealed it in an envelope to maintain confidentiality. Each student had a self-generated identification code derived from their birth month, name (e.g. second letter of first name) and mother's name, allowing student data to be de-identified while also being linked across multiple years. Detail on the longitudinal data linkage procedure is available in a technical report.[28]

Measures

School variables

School-level data used in this study were collected in Y_1 or Y_2 (whichever year the school first participated in COMPASS) using the COMPASS School Policies and Practices Questionnaire. The COMPASS staff member also collected school handbooks and written policies. School location (urban or rural) and neighbourhood socioeconomic status (SES) were determined using the postal code and 2011 Canadian census data. School-level characteristics included in this study are listed in Table 1, in addition to the presence of

TABLE 1
Characteristics of schools participating in both Year 2 (2013/14) and Year 4 (2015/16) of the COMPASS study in Ontario and Alberta, Canada

		N = 78	
		n	%
Demographic			
Enrolment	1–500 students	32	41.0
	501–1000 students	39	50.0
	≥ 1001 students	7	9.0
SES[a] (CAD)	25 000–50 000	7	9.0
	50 001–75 000	51	65.4
	75 001–100 000	17	21.8
	≥ 100 001	3	3.9
Location	Rural	1	1.3
	Small urban	34	43.6
	Medium urban	13	16.7
	Large urban	30	38.5
Social environment			
School priority of PA	1st–3rd	17	21.8
	4th–6th	35	44.9
	7th–10th	18	23.1
	Missing	8	10.3
Promotion of PA events	Yes	69	88.5
	No	8	10.3
	Missing	1	1.3
Physical environment			
Curtains for changing	Both girls and boys	21	26.9
	Girls only	15	19.2
	Boys only	1	1.3
	None	34	43.6
	Missing	7	9.0
Secure lockers in change room	Yes	62	79.5
	No	15	19.2
	Missing	1	1.3
Showers	Both girls and boys	69	88.5
	Girls only	0	0.0
	Boys only	2	2.6
	None	4	5.1
	Missing	3	3.9
Partnerships and services			
Organizations providing support (check all that apply)	Local public health	37	52.6
	Nongovernmental organization	47	60.3
	Parks and recreation	21	27.0
	Youth organizations	23	29.5
	Health and fitness club	39	50.0
	Consultant/specialist	27	34.6
Teaching and learning			
Noncurricular physical activity programs	Both intramurals and non-competitive clubs	35	44.9
	Intramural only	15	19.2
	Non-competitive only	15	19.2
	None	13	16.7
Healthy school policy			
Written policy	Yes	45	57.7
	No	24	30.8
	Missing	9	11.5
Data from student health assessment used to plan	Yes	30	38.5
	No	48	61.5

Continued on the following page

indoor facilities, outdoor facilities, gymnasium, change rooms and interschool or varsity sports teams.

Student variables

Student demographics from the annual COMPASS student questionnaire[29] included in this study were gender, grade and ethnicity.

The outcome variable was student achievement of all three PA recommendations, consistent with previous research.[30] Three questions from the COMPASS questionnaire were used to determine this status: (1) "Mark how many minutes of HARD physical activities you did on each of the last 7 days"; (2) "Mark how many minutes of MODERATE physical activities you did on each of the last 7 days"; and (3) "On how many days in the last 7 days did you do exercises to strengthen or tone your muscles?" (with examples provided). The first two questions were used to determine whether the student achieved an average of 60 minutes or more of MVPA per day. Students who participated in any minutes of hard PA (i.e. greater than 0 minutes) at least three days per week were classified as achieving the VPA recommendation.

A validation study found that while these two questions individually had low validity when compared to objectively measured PA (Pearson r = 0.21 and 0.27, respectively), the validity was higher when combined (r = 0.31) and was similar to other self-report questionnaires.[31] Test-retest reliability for these two questions was moderate (r = 0.69 and 0.57, respectively, and r = 0.68 combined). There is no validity or reliability data available for the muscle and bone strengthening activity question, but since there is no intensity needed or minimum number of minutes, it is expected that students could estimate the number of days.

Students who achieved an average of 60 minutes or more per day of MVPA, VPA three or more days per week and muscle and bone strengthening activity three or more days per week were designated as achieving all three PA recommendations; all others were classified as not achieving the full recommendations (even if they achieved one or two of the recommendations).

Statistical analysis

Descriptive analysis was conducted for school- and student-level characteristics.

TABLE 1 (continued)
Characteristics of schools participating in both Year 2 (2013/14) and Year 4 (2015/16) of the COMPASS study in Ontario and Alberta, Canada

		N = 78	
		n	%
Access during non-instructional time:			
Indoor facilities	Yes	53	68.0
	No	25	32.1
Outdoor facilities	Yes	67	85.9
	No	9	11.5
	Missing	2	2.6
Equipment	Always	23	29.5
	Sometimes	46	59.0
	Never	9	11.5
Access outside of school hours:			
Gymnasium	Yes	61	78.2
	No	17	21.8
Indoor facilities	Yes	60	76.9
	No	17	21.8
	Missing	1	1.3
Outdoor facilities	Yes	66	84.6
	No	11	14.1
	Missing	1	1.3
Equipment	Yes	54	69.2
	No	23	29.5
	Missing	1	1.3
School board provided resource:			
Staff time	Yes	46	59.0
	No	27	34.6
	Missing	5	6.4
Additional space	Yes	25	32.1
	No	47	60.3
	Missing	6	7.7
Budget to improve health (CAD)	≥ 1001	27	34.6
	1– 1000	12	15.4
	No funding	31	39.7
	Missing	8	10.3

Abbreviations: CAD, Canadian dollars; PA, physical activity; SES, socioeconomic status.

Notes: Due to rounding, the total frequencies for some variables do not sum to 100% exactly.

[a] SES is the neighbourhood median household income based on the school's postal code and data from the 2011 Canadian census.

Since missing data at the school-level were considered to be missing not at random (i.e. the likelihood of missingness is based on the unobserved data itself), the missing indicator approach was used to handle missing data.[32] For each PA recommendation and the full set of three recommendations, comparisons across genders were performed using chi-square tests. Cross-tabulations were used to determine the proportion of students meeting the recommendations in Y_2 and who continued meeting them in Y_4, as well as the proportion of students not meeting the recommendations in Y_2 who were able to meet them in Y_4.

The regression analyses were stratified by gender, and then further stratified into two groups based on whether the student was achieving all three recommendations in Y_2, for a total of four strata. Due to the hierarchical nature of the data, multilevel logistic regression models (random intercept only, clustered by school) were conducted for each of the four student strata (gender and baseline PA status), with the binary outcome of achieving all three PA recommendations in Y_4 or not. For each stratum, a null model was run to determine variability across schools in the outcome. To calculate the intraclass correlation (ICC), since the outcome was dichotomous, the student-level portion of the error variance was set at $\pi^2/3$, which is approximately 3.29.[33]

Next, full models were run for each student stratum, with all school characteristics as Level 2 variables (excluding those with a prevalence of 95% or higher due to lack of heterogeneity) and student grade and ethnicity added as Level 1 variables (potential student-level confounders). The four full models were assessed for multicollinearity (e.g. high-variance inflation factors). All analyses were conducted using RStudio version 1.3.1 (RStudio Team, Boston, MA, USA) and R version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). The mixed_model procedure of the GLMMadaptive package was used with binomial distribution, logit link and 100 expectation-maximization iterations for each regression. As this was an exploratory study, the results and discussion are focussed primarily on the statistically significant findings.

Results

Seventy-eight schools participated in both Y_2 and Y_4 of COMPASS; 10 schools from Y_2 had dropped out by Y_4 (nine from Ontario and one from Alberta), primarily due to labour disputes in Ontario.[27] The prevalence of PA-related school characteristics within the CSH framework are provided in Table 1 and demonstrate heterogeneity across schools. In addition, 98.7% (n = 77) had a gymnasium, 97.4% (n = 76) had other indoor facilities, 94.9% (n = 74) had change rooms, 98.7% (n = 77) had outdoor facilities, and 100% (n = 78) had interschool or varsity sports.

Y_2 physical activity levels

Within these schools, 10 160 students in Grades 9 and 10 participated in COMPASS in both Y_2 and Y_4, and 9870 students (97.1%) reported their PA in both years. In Y_2, only 47.7% (n = 4706) of students reported achieving all three PA recommendations; 84.6% (n = 8351) reported achieving the VPA recommendation, 77.8% (n = 7683) the MVPA recommendation, and 55.0% (n = 5430) the muscle and bone strengthening activity recommendation. A significantly higher proportion of male students compared to female students reported achieving each of the separate PA recommendations as well as the full set of three recommendations ($p < 0.001$; Table 2).

Proportion achieving the PA recommendations over time

Of the 4706 students achieving all three recommendations in Y_2, 56.9% (n = 2679)

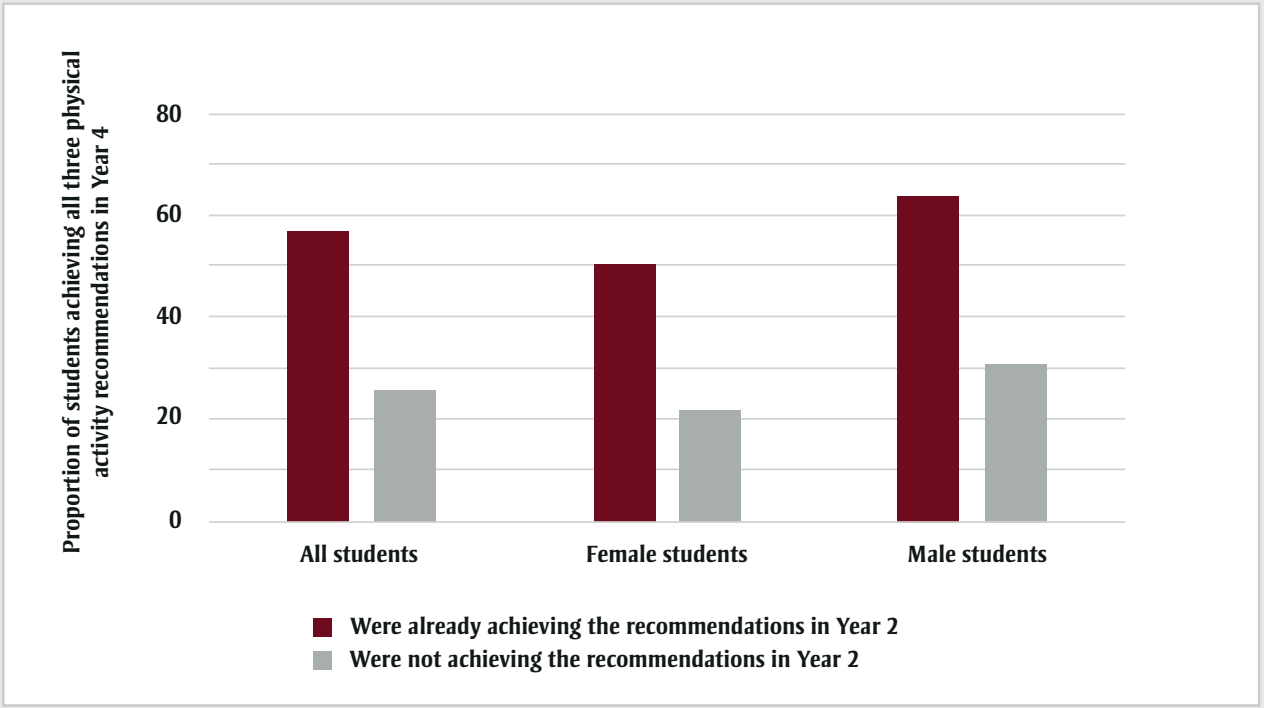
TABLE 2
Physical activity behaviour of students, by gender, attending schools that participated in Year 2 (2013/14) and Year 4 (2015/16) of the COMPASS study in Ontario and Alberta, Canada

	Total (n = 9870)		Female (n = 5200)		Male (n = 4670)		Chi-square test	
	n	%	n	%	n	%	Difference in male compared to female, % (95% CI)	*p*-value
Grade in Y_2								
9	5309	53.8	2722	52.4	2587	55.4	—	—
10	4561	46.2	2478	47.7	2083	44.6	—	—
Ethnicity								
White only	7579	76.8	4046	77.8	3533	75.7	—	—
Other	2291	23.2	1154	22.2	1137	24.4	—	—
Physical activity in Y_2								
Meets MVPA recommendation	7683	77.8	3830	73.7	3853	82.5	8.9 (7.2–10.5)	< 0.001
Meets VPA recommendation	8351	84.6	4281	82.3	4070	87.2	4.8 (3.4–6.2)	< 0.001
Meets MBSA recommendation	5430	55.0	2736	52.6	2694	57.7	5.1 (3.1–7.0)	< 0.001
Achieves all three recommendations	4706	47.7	2279	43.8	2427	52.0	8.1 (6.2–10.1)	< 0.001

Abbreviations: CI, confidence interval; MBSA, muscle and bone strengthening activity; MVPA, moderate-to-vigorous physical activity; VPA, vigorous physical activity; Y_2, Year 2 of the COMPASS cohort.

continued achieving them in Y_4, and of the 5164 students not achieving all three recommendations in Y_2, 25.6% (n = 1321) were able to achieve them by Y_4 (Figure 1). Among the 2279 female students who were achieving all three recommendations in Y_2, only 49.9% (n = 1138) continued to achieve them in Y_4; of the 2921 who were not achieving the recommendations in Y_2, 21.6% (n = 632) improved to achieve them in Y_4. Finally, for the 2427 male students who were achieving all three recommendations in Y_2, 63.5% (n = 1541) were still achieving the recommendations in Y_4, and of the 2243 male students not achieving the recommendations in Y_2, 30.7% (n = 689) were able to achieve them by Y_4.

FIGURE 1
Proportion of students achieving all three physical activity recommendations[a] in Year 4 (2015/16) of the COMPASS study in Ontario and Alberta, Canada, by sex and by baseline physical activity status in Year 2 (2013/14)



[a] The three national physical activity recommendations are: (1) an average of 60 minutes or more of moderate-to-vigorous physical activity daily; (2) vigorous physical activity at least three days per week; and (3) muscle and bone strengthening activity at least three days per week.

School characteristics associated with achieving the physical activity recommendations in Year 4

The ICC was very low across all four models (ICC = 0.007–0.027) but was highest among females achieving the PA recommendations in Y_2 (Table 3).

Social environment

Female students not achieving the PA recommendations in Y_2 had significantly lower odds of achieving them in Y_4 if their school gave PA a priority ranking of 7th to 10th in Y_2 compared to a priority of 1st to 3rd (adjusted odds ratio [AOR] = 0.56; 95% CI: 0.32–0.97). This group also had lower odds of achieving the recommendations if their school promoted PA events (AOR = 0.62; 95% CI: 0.40–0.94). In contrast, male students not achieving the guidelines in Y_2 had higher odds of achieving them in Y_4 if their school ranked PA in the lowest priority group (AOR = 1.85; 95% CI: 1.09–3.14) and promoted PA events (AOR = 1.53; 95% CI: 1.00–2.33). For students already achieving the recommendations in Y_2, no significant associations were found between the social environment factors in this study and achieving the recommendations in Y_4.

Physical environment

Male students achieving the recommendations in Y_2 had lower odds of achieving them in Y_4 if their school provided curtains in the change room (AOR = 0.67; 95% CI: 0.50–0.89). Female students who were achieving the recommendations had lower odds of achieving them in Y_4 if their school provided showers (AOR = 0.62; 95% CI: 0.38–0.99). While not significant, it is notable that both female and male students who were not achieving the recommendations in Y_2 were also trending towards lower odds of achieving them in Y_4 if their school provided showers (AOR = 0.63; 95% CI: 0.39–1.03 and AOR = 0.62; 95% CI: 0.38–1.01, respectively).

Partnerships and services

Male students achieving the recommendations in Y_2 had greater odds of continuing to meet them in Y_4 if their school was partnered with public health (AOR = 1.37; 95% CI: 1.05–1.79).

Teaching and learning

There were no significant associations observed between schools offering non-competitive or intramural PA opportunities and student odds of achieving the PA recommendations in Y_4 for any of the student strata.

Policy

Looking first at policies for student access during non-instructional time within school hours, female students not achieving the recommendations in Y_2 had significantly lower odds achieving them in Y_4 if their school provided access to indoor facilities (AOR = 0.64; 95% CI: 0.45–0.93), whereas male students achieving the recommendations in Y_2 had higher odds of continuing to meet them if their school provided access to equipment sometimes or always (AOR = 1.51; 95% CI: 1.00–2.27 and AOR = 1.56; 95% CI: 1.09–2.22, respectively). For after-school hours, male students not achieving the recommendations in Y_2 had lower odds of achieving them in Y_4 if their school provided access to the gymnasium (AOR = 0.50; 95% CI: 0.24–0.86). Finally, considering resources provided by the school board to improve student health, female students not achieving the recommendations in Y_2 had higher odds of achieving them in Y_4 if their school received staff time to support student health, but lower odds if their school was provided additional space (AOR = 1.42; 95% CI: 1.02–1.98 and AOR = 0.62; 95% CI: 0.39–0.97, respectively).

Discussion

Despite finding that the majority of students in this sample were not achieving all three PA recommendations, some school characteristics within the CSH framework were positively associated with students achieving these recommendations after two years. For students not achieving the recommendations in Y_2, one-quarter achieved them two years later; this is a meaningful achievement given the evidence that PA typically declines with age.[3,7]

Among students achieving the recommendations in Y_2, nearly half were no longer attaining this standard after two years, demonstrating a need not only to focus on increasing PA among insufficiently active students, but also to support active students to continue being active. There was a negligible clustering effect for male students already achieving the recommendations in Y_2, meaning that male students in this stratum sampled from the same school were just as similar to each other as they were to male students from other schools. While the other student strata also had low ICCs (2%–3%), at a population level, small shifts in behaviour can have a large impact,[34] and there were school characteristics associated with PA behaviours over time for all four student groups studied.

For the first component of CSH, social and physical environment, the social environment was associated with achieving the PA recommendations only for students not already achieving them at baseline in our sample. Specifically, a school's low prioritization of PA relative to other health-related issues was positively associated with male students achieving the guidelines two years later but negatively associated for female students. This contrasting finding may suggest that school administrators are prioritizing PA based on observed male engagement with PA programming and not female engagement. Another potential explanation is that the health-related issues being prioritized over PA (e.g. tobacco use, cannabis use, bullying) may indirectly increase PA among males more than females.[35-37] Another contrasting finding was also observed for the school's promotion of PA events, whereby males attending schools that promoted PA events had higher odds of achieving the recommendations after two years, but female students had lower odds. The consistency in these contrasting male and female results further supports that male PA levels may be driving school prioritization and the types of PA events being promoted, but additional research is needed to better understand the dynamics behind these observed results.

For the physical environment, we unexpectedly found that providing curtains and showers was negatively associated with achieving the guidelines for some student groups. In fact, the provision of showers was the only school-level factor significantly associated with achieving the guidelines for female students already achieving the recommendations in Y_2, and it was a negative relationship. Although the association was borderline statistically significant (p = 0.046), the observed odds ratios and confidence intervals were nearly identical across three of the four student strata, which pragmatically suggests there may be a pattern that warrants further examination. There is insufficient information within this study to know the degree of privacy offered in the shower area, but previous research has found that social pressures and psychological

TABLE 3
Adjusted odds ratio of achieving PA guidelines in Year 4 (2015/16) based on status in Year 2 (2013/14) and according to student and school characteristics, from the COMPASS study in Ontario and Alberta, Canada

	Adjusted odds of students achieving all 3 PA recommendations[a] in Y_4 for those who were already achieving them at baseline (Y_2)		Adjusted odds of students achieving all 3 PA recommendations in Y_4 for those who were <u>not</u> achieving them at baseline (Y_2)	
	Females n = 2279	Males n = 2427	Females n = 2921	Males n = 2243
	AOR[b] (95% CI)	AOR[b] (95% CI)	AOR[b] (95% CI)	AOR[b] (95% CI)
Intraclass correlation (null model)	0.03	0.01	0.02	0.02
Level 1: Student-level				
Grade in Y_2				
Grade 10 (ref = Grade 9)	**0.72 (0.61–0.85)*****	0.92 (0.78–1.10)	**0.80 (0.67–0.96)***	**0.81 (0.67–0.97)***
Ethnicity				
Other (ref = White only)	1.17 (0.94–1.45)	**1.35 (1.10–1.64)****	1.06 (0.85–1.32)	0.88 (0.71–1.10)
Level 2: School-level				
Social environment				
Enrolment				
1–500 students	1.00	1.00	1.00	1.00
501–1000 students	0.92 (0.67–1.28)	1.12 (0.82–1.55)	1.12 (0.79–1.57)	1.03 (0.72–1.46)
≥ 1001 students	0.69 (0.36–1.41)	1.40 (0.71–2.76)	1.30 (0.62–2.73)	0.97 (0.47–2.04)
SES[c] (CAD)				
25 000–50 000	1.00	1.00	1.00	1.00
50 001–75 000	1.21 (0.77–1.91)	0.81 (0.52–1.26)	0.96 (0.60–1.55)	1.36 (0.83–2.22)
75 001–100 000	0.95 (0.55–1.63)	0.70 (0.42–1.18)	0.96 (0.56–1.74)	0.99 (0.56–1.75)
≥ 100 001	1.87 (0.89–3.97)	1.06 (0.50–2.23)	1.25 (0.55–2.84)	2.06 (0.85–4.96)
School priority of PA				
1st–3rd	1.00	1.00	1.00	1.00
4th–6th	0.95 (0.67–1.35)	0.84 (0.59–1.18)	0.74 (0.52–1.05)	1.30 (0.89–1.89)
7th–10th	0.88 (0.53–1.46)	1.04 (0.62–1.73)	**0.56 (0.32–0.97)***	**1.85 (1.09–3.14)***
Missing	1.25 (0.66–2.38)	0.63 (0.35–1.15)	0.72 (0.38–1.39)	1.73 (0.91–3.29)
Promotes PA events	0.97 (0.64–1.45)	1.01 (0.68–1.49)	**0.62 (0.40–0.94)***	**1.53 (1.00–2.33)***
Physical environment				
Location				
Rural/small urban	1.00	1.00	1.00	1.00
Medium urban	1.01 (0.68–1.50)	1.26 (0.84–1.89)	0.79 (0.52–1.18)	0.97 (0.62–1.52)
Large urban	1.21 (0.78–1.88)	1.23 (0.80–1.88)	0.67 (0.41–1.07)	0.96 (0.59–1.54)
Curtains available	0.99 (0.73–1.33)	**0.67 (0.50–0.89)****	1.13 (0.83–1.54)	1.12 (0.83–1.50)
Secure lockers available	0.86 (0.54–1.36)	1.24 (0.80–1.94)	1.37 (0.84–2.24)	0.88 (0.53–1.44)
Showers available	**0.62 (0.38–0.99)***	1.09 (0.69–1.73)	0.63 (0.39–1.03)	0.62 (0.38–1.01)
Partnerships and services				
Public health	1.13 (0.84–1.51)	**1.37 (1.05–1.79)***	0.80 (0.59–1.10)	1.06 (0.77–1.45)
Nongovernmental organization	0.80 (0.63–1.03)	0.96 (0.75–1.24)	0.92 (0.69–1.22)	0.82 (0.63–1.08)
Parks and recreation	0.96 (0.70–1.31)	0.88 (0.64–1.21)	1.33 (0.96–1.85)	0.85 (0.60–1.21)
Youth organizations	1.24 (0.89–1.72)	1.38 (0.97–1.94)	1.04 (0.73–1.48)	1.07 (0.75–1.53)
Health and fitness club	1.07 (0.77–1.47)	0.97 (0.70–1.33)	1.22 (0.86–1.73)	1.21 (0.86–1.70)
Consultant/specialist	0.93 (0.68–1.27)	1.18 (0.86–1.62)	0.99 (0.72–1.37)	1.14 (0.82–1.58)

Continued on the following page

TABLE 3 (continued)
Adjusted odds ratio of achieving PA guidelines in Year 4 (2015/16) based on status in Year 2 (2013/14) and according to student and school characteristics, from the COMPASS study in Ontario and Alberta, Canada

	Adjusted odds of students achieving all 3 PA recommendations[a] in Y_4 for those who were already achieving them at baseline (Y_2)		Adjusted odds of students achieving all 3 PA recommendations in Y_4 for those who were <u>not</u> achieving them at baseline (Y_2)	
	Females n = 2279	Males n = 2427	Females n = 2921	Males n = 2243
	AOR[b] (95% CI)	AOR[b] (95% CI)	AOR[b] (95% CI)	AOR[b] (95% CI)
Teaching and learning				
Noncurricular PA programs				
Intramural and non-competitive	0.80 (0.55–1.17)	1.07 (0.73–1.58)	1.07 (0.72–1.58)	1.29 (0.85–1.97)
Intramural only	0.78 (0.51–1.18)	1.19 (0.77–1.84)	1.30 (0.82–2.06)	0.90 (0.56–1.47)
Non-competitive only	0.95 (0.60–1.51)	0.84 (0.53–1.33)	1.13 (0.70–1.83)	1.32 (0.80–2.19)
None	1.00	1.00	1.00	1.00
Healthy school policy				
Has written policy	0.82 (0.56–1.19)	0.99 (0.71–1.38)	0.82 (0.55–1.21)	0.85 (0.59–1.23)
Uses data to plan	1.14 (0.85–1.53)	1.02 (0.76–1.38)	0.73 (0.53–1.02)	0.97 (0.71–1.33)
Access during non-instructional time:				
Indoor facility	1.11 (0.78–1.57)	0.83 (0.59–1.17)	**0.64 (0.45–0.93)***	1.06 (0.72–1.57)
Outdoor facility	0.75 (0.49–1.15)	0.77 (0.49–1.19)	0.78 (0.51–1.20)	0.84 (0.53–1.36)
Equipment				
Always	1.20 (0.83–1.73)	**1.56 (1.09–2.22)***	1.40 (0.94–2.09)	1.01 (0.68–1.51)
Sometimes	1.04 (0.65–1.66)	**1.51 (1.00–2.27)***	1.51 (0.94–2.42)	0.93 (0.60–1.45)
Never	1.00	1.00	1.00	1.00
Access after school hours:				
Gymnasium	0.72 (0.41–1.26)	0.90 (0.55–1.47)	0.79 (0.44–1.45)	**0.50 (0.29–0.86)***
Indoor facility	1.26 (0.88–1.81)	1.01 (0.71–1.45)	1.26 (0.84–1.89)	1.38 (0.90–2.10)
Outdoor facility	1.26 (0.87–1.83)	1.04 (0.71–1.54)	1.30 (0.87–1.95)	1.01 (0.66–1.54)
Equipment	0.75 (0.49–1.15)	1.16 (0.78–1.71)	0.68 (0.42–1.09)	0.95 (0.61–1.48)
Resources from school board:				
Staff time	0.82 (0.60–1.11)	0.88 (0.66–1.18)	**1.42 (1.02–1.98)***	1.00 (0.71–1.41)
Space	0.98 (0.64–1.51)	1.11 (0.75–1.65)	**0.62 (0.39–0.97)***	1.11 (0.70–1.76)
Budget (CAD)				
No funding	1.00	1.00	1.00	1.00
1–1000	1.15 (0.80–1.64)	1.29 (0.87–1.92)	1.29 (0.89–1.89)	0.84 (0.56–1.27)
≥ 1001	1.00 (0.75–1.33)	0.92 (0.71–1.21)	1.25 (0.92–1.70)	0.81 (0.60–1.10)
Missing	0.76 (0.47–1.22)	1.01 (0.64–1.61)	1.17 (0.72–1.91)	0.86 (0.52–1.42)

Abbreviations: AOR, adjusted odds ratio; CAD, Canadian dollars; PA, physical activity; ref, reference category; SES, socioeconomic status; Y_2, Year 2 of the COMPASS cohort; Y_4, Year 4 of the COMPASS cohort.

Notes: Unless otherwise stated, the reference category is any response other than a definitive "Yes" (e.g. no, not applicable, no response, uncodable). Bolded data are statistically significant.

[a] The three national physical activity recommendations are: (1) an average of 60 minutes or more of moderate-to-vigorous physical activity daily; (2) vigorous physical activity at least three days per week; and (3) muscle and bone strengthening activity at least three days per week.

[b] AORs are adjusted for all student and school variables presented within the table.

[c] SES is the neighbourhood median household income based on the school's postal code and data from the 2011 Canadian census.

*$p < 0.05$
**$p < 0.01$
***$p < 0.001$

discomfort in using showers at school may contribute to students shying away from PA as they progress through puberty.[38-40]

In addition, male students achieving the recommendations at baseline had lower odds of continuing to achieve them if their school provided privacy curtains in the change room. An in-depth examination is needed into the specific features of the showers and curtains being provided (e.g. degree of privacy) and the social dynamics around their use. For example, there may be bullying or stigma attached to boys who choose to use the curtains or who avoid the showers. Previous studies have shown that boys who withdraw from physical education classes when they are no longer mandatory often make this decision based on experiencing bullying and abuse by peers, not due to a disinterest in PA.[37]

For the second component of CSH, partnerships and services, the results of this study suggest that a public health partnership can be beneficial in supporting active male students in maintaining their PA over time; however, it also suggests that public health materials or other supports being provided may need to be expanded to target other student groups (i.e. female students and male students not already achieving the guidelines). This relationship between partnerships and PA had been reported in a study of middle schools,[19] but has not been explored in a secondary school sample. This is an important finding, as linking schools with local public health units is both an affordable and a feasible intervention that could be promoted as part of CSH for schools seeking to improve student PA.

For the teaching and learning component of CSH, there were no associations found between providing intramural or noncompetitive PA options and students achieving the PA recommendations after two years. This was unexpected, since previous research found that female students were more likely to participate in PA when provided these options.[40,41] It may be that these PA offerings only facilitate achieving one or two of the PA recommendations and not all three. While beyond the scope of the work presented here, additional investigation with COMPASS data should explore this hypothesis.

For the policy component of CSH, male students already meeting the recommendations had approximately 50% higher odds of continuing to meet them if their school permitted access to equipment during non-instructional time. However, this relationship was not observed during the after-school period, indicating that the period in which students are permitted access is important for having the desired effect. Receiving staff time for health from the school board was significantly associated with female students transitioning from not achieving the PA recommendations to achieving them two years later. Since it is well-established that female students are significantly less active than males and tend to stop being active at a younger age,[3,7] interventions associated with increasing PA among this at-risk group are highly desirable. Further research into how this additional staff time is being used in practice could inform similar approaches for other schools.

Strengths and limitations

Two major strengths of this study were the large cohort of students with linked PA data over time and having data collected simultaneously from the schools they were attending. This unique resource available through COMPASS allowed for a longitudinal analysis of the associations between student PA and many school characteristics within the CSH framework. Another strength was that this exploratory study analyzed many school-level factors simultaneously, which allowed for the relative association of each school-level characteristic with student PA to be compared within the models. This approach has laid the foundation for future research studies. Although this initial study examined a composite outcome of achieving all three PA recommendations, future studies may examine the school characteristics associated with students achieving each type of PA independently (e.g. VPA, muscle and bone strengthening exercises). Also, it was beyond the scope of this study to examine interactions between the school-level variables, but the complexity of these relationships could be examined in future research; for example, whether the association between intramural PA opportunities and student PA is moderated by school SES. Finally, future studies may expand to include schools from other provinces and follow students over a longer period of time.

There were some limitations with this study. First, the self-reported PA data is a limitation, since there can be recall bias and it is known that students tend to underestimate their moderate PA and overestimate their VPA.[31] However, this is expected to be partially mitigated by the longitudinal nature of the study, since individuals may have a similar degree of self-report bias at both time points and were being compared against their own baseline. Second, there may have been nonresponse bias in the students who did not participate or did not provide complete data; however, there was a high participation rate in COMPASS each year (79%–80%) and missing PA data in this sample was less than 2%. Third, the follow-up period for this longitudinal study was relatively short, which could have produced some of the unexpected findings (e.g. negative association between providing indoor space and achieving the PA recommendations). Fourth, the schools studied were from Ontario and Alberta, and the generalizability to schools from other provinces may be limited due to differences in provincial-level policies (e.g. mandatory physical education credits) or grade-level divisions. Fifth, there were limitations with the school-level data available. There were missing data for some school school-level variables, which was handled using the missing indicator approach,[32] but using a combined multiple imputation and missing indicator approach may be a better strategy for data missing not at random[42] and should be considered in future work. Also, only the baseline school characteristics were included in this study, since there were inadequate data on school-level changes over time for all factors included in the models. Future studies are needed that can incorporate these school-level changes into the longitudinal model.

Conclusion

In conclusion, schools can play a role in supporting students in both achieving the PA recommendations and maintaining them over time, countering the well-documented decline in PA behaviour during this life stage. The significant school-based factors identified in this study were generally affordable, feasible changes (e.g. public health partnership, access to equipment during school hours) that are already being implemented by other schools. The CSH framework can guide schools in providing a health

promoting environment for students, but the elements that will be most effective depend on the student subgroup being targeted and the context of the schools themselves.

Acknowledgements

This work was supported by a bridge grant from the Canadian Institutes of Health Research (CIHR) Institute of Nutrition, Metabolism and Diabetes through the "Obesity—Interventions to Prevent or Treat" priority funding awards (OOP-110788; to STL) and an operating grant from the CIHR Institute of Population and Public Health (MOP-114875; to S.T.L). VC is supported by a CIHR New Investigator Salary Award and a Killam Accelerator Research Award.

Conflicts of interest

Scott Leatherdale is an Associate Scientific Editor with the HPCDP Journal, but has recused himself from the review process for this paper. The authors declare there are no other conflicts of interest.

Authors' contributions and statement

MP and SL conceptualized the work; SL and VC led the data acquisition; MP conducted the analysis and all coauthors contributed to the interpretation of the data; MP drafted the paper; all authors contributed to revisions and approved the final 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.

References

1. Tremblay MS, Carson V, Chaput JP, et al. Canadian 24-hour movement guidelines for children and youth: an integration of physical activity, sedentary behaviour, and sleep. Appl Physiol Nutr Metab. 2016;41(6):S311-S327. https://doi.org/10.1139/apnm-2016-0151

2. Strong WB, Malina RM, Blimkie CJR, et al. Evidence based physical activity for school-age youth. J Pediatr. 2005; 146(6):732-7. https://doi.org/10.1016/j.jpeds.2005.01.055

3. Statistics Canada. Table 13-10-0388-01, "Household population meeting/not meeting the Canadian physical activity guidelines" [Internet]. Ottawa (ON): Government of Canada; 2020 [cited 2020 Aug 13]. Available from: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310038801

4. Brown JC, Winters-Stone K, Lee A, Schmitz KH. Cancer, physical activity, and exercise. Compr Physiol. 2012; 2(4):2775-809. https://doi.org/10.1002/cphy.c120005

5. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ. 2006;174(6):801–9. https://doi.org/10.1503/cmaj.051351

6. Malina RM. Physical activity and fitness: pathways from childhood to adulthood. Am J Hum Biol. 2001; 13(2):162-72. https://doi.org/10.1002/1520-6300(200102/03)13:2<162::AID-AJHB1025>3.0.CO;2-T

7. Roberts KC, Yao X, Carson V, Chaput J-P, Janssen I. Meeting the *Canadian 24-Hour Movement Guidelines for Children and Youth*. Health Rep. 2017;28(10):3-7.

8. Telama R, Yang X, Viikari J, Välimäki I, Wanne O, Raitakari O. Physical activity from childhood to adulthood. Am J Prev Med. 2005;28(3):267-73. https://doi.org/10.1016/j.amepre.2004.12.003

9. Bronfenbrenner U. Toward an experimental ecology of human development. Am Psychol. 1977;32(7):513-31. https://doi.org/10.1037/0003-066X.32.7.513

10. Rainham DG, Bates CJ, Blanchard CM, Dummer TJ, Kirk SF, Shearer CL. Spatial classification of youth physical activity patterns. Am J Prev Med. 2012;42(5):e87-e96. https://doi.org/10.1016/j.amepre.2012.02.011

11. Kriemler S, Meyer U, Martin E, van Sluijs EM, Andersen LB, Martin BW. Effect of school-based interventions on physical activity and fitness in children and adolescents: a review of reviews and systematic update. Br J Sports Med. 2011;45(11):923-30. https://doi.org/10.1136/bjsports-2011-090186

12. Morton KL, Atkin AJ, Corder K, Suhrcke M, van Sluijs EM. The school environment and adolescent physical activity and sedentary behaviour: a mixed-studies systematic review. Obes Rev. 2016;17(2):142-58. https://doi.org/10.1111/obr.12352

13. Dobbins M, Husson H, Decorby K, LaRocca RL. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. Cochrane Database Syst Rev. 2013;2013(2): CD007651. https://doi.org/10.1002/14651858.CD007651.pub2

14. Leatherdale ST. Natural experiment methodology for research: a review of how different methods can support real-world research. Int J Soc Res Methodol. 2018;22(1):19-35. https://doi.org/10.1080/13645579.2018.1488449

15. Veugelers PJ, Schwartz ME. Comprehensive school health in Canada. Can J Public Health. 2010;101(Suppl 2):S5-S8. https://doi.org/10.1007/BF03405617

16. Joint Consortium for School Health (JCSH). What is Comprehensive School Health? [Internet]. Charlottetown (PE): JCSH; 2016 [cited 2021 Jun 19]. Available from: http://www.jcsh-cces.ca/en/concepts/comprehensive-school-health/

17. Hohepa M, Scragg R, Schofield G, Kolt GS, Schaaf D. Social support for youth physical activity: importance of siblings, parents, friends and school support across a segmented school day. Int J Behav Nutr Phys Act. 2007; 4:54. https://doi.org/10.1186/1479-5868-4-54

18. Pate RR, Ward DS, Saunders RP, Felton G, Dishman RK, Dowda M. Promotion of physical activity among high-school girls: a randomized controlled trial. Am J Public Health. 2005;95(9):1582-7. https://doi.org/10.2105/AJPH.2004.045807

19. Leatherdale ST, Manske S, Faulkner G, Arbour K, Bredin C. A multi-level examination of school programs, policies and resources associated with physical activity among elementary school youth in the PLAY-ON study. Int J Behav Nutr Phys Act. 2010;7(1):6. https://doi.org/10.1186/1479-5868-7-6

20. Pabayo R, O'Loughlin J, Gauvin L, Paradis G, Gray-Donald K. Effect of a ban on extracurricular sports activities by secondary school teachers on physical activity levels of adolescents: a multilevel analysis. Health Educ Behav. 2006;33(5):690-702. https://doi.org/10.1177/1090198105285327

21. Nichol ME, Pickett W, Janssen I. Associations between school recreational environments and physical activity. J Sch Health. 2009;79(6):247-54. https://doi.org/10.1111/j.1746-1561.2009.00406.x

22. Naylor P-J, Macdonald HM, Zebedee JA, Reed KE, McKay HA. Lessons learned from Action Schools! BC—an 'active school' model to promote physical activity in elementary schools. J Sci Med Sport. 2006;9(5):413-23. https://doi.org/10.1016/j.jsams.2006.06.013

23. Hobin EP, Leatherdale ST, Manske S, Dubin JA, Elliott S, Veugelers P. A multilevel examination of gender differences in the association between features of the school environment and physical activity among a sample of grades 9 to 12 students in Ontario, Canada. BMC Public Health. 2012; 12:74. https://doi.org/10.1186/1471-2458-12-74

24. Leatherdale ST, Brown KS, Carson V, et al. The COMPASS study: a longitudinal hierarchical research platform for evaluating natural experiments related to changes in school-level programs, policies and built environment resources. BMC Public Health. 2014; 14:331. https://doi.org/10.1186/1471-2458-14-331

25. Thompson-Haile A, Leatherdale ST. Baseline sampling and recruitment results [Internet]. [COMPASS Technical Report Series.] Waterloo (ON): University of Waterloo; 2013 [cited 2022 Jan 18]. Available from: https://uwaterloo.ca/compass-system/publications/baseline-sampling-and-recruitment-results

26. Wagner M, Bredin C, Thompson-Haile A, Leatherdale ST. Alberta baseline sampling and recruitment results [Internet]. [COMPASS Technical Report Series.] Waterloo (ON): University of Waterloo; 2015 [cited 2022 Jan 18]. Available from: https://uwaterloo.ca/compass-system/publications/alberta-baseline-sampling-and-recruitment-results

27. Reel R, Bredin C, Battista K, Leatherdale ST. COMPASS year 3 and 4 school recruitment and retention [Internet]. [COMPASS Technical Report Series.] Waterloo (ON): University of Waterloo; 2017 [cited 2022 Jan 18]. Available from: https://uwaterloo.ca/compass-system/compass-year-3-and-4-school-recruitment-and-retention

28. Bredin C, Leatherdale ST. Methods for linking COMPASS student-level data over time [Internet]. [COMPASS Technical Report Series.] Waterloo (ON): University of Waterloo; 2013 [cited 2021 Aug 18]. Available from: https://uwaterloo.ca/compass-system/publications/methods-linking-compass-student-level-data-over-time

29. Bredin C, Leatherdale S. Development of the COMPASS student questionnaire [Internet]. [COMPASS Technical Report Series.] Waterloo (ON): University of Waterloo; 2014 [cited 2021 Aug 18]. Available from: https://uwaterloo.ca/compass-system/publications/development-compass-student-questionnaire

30. Harvey A, Faulkner G, Giangregorio L, Leatherdale ST. An examination of school- and student-level characteristics associated with the likelihood of students' meeting the Canadian physical activity guidelines in the COMPASS study. Can J Public Health. 2017; 108(4):e348-e354. https://doi.org/10.17269/cjph.108.5925

31. Leatherdale ST, Laxer RE, Faulkner G. Reliability and validity of the physical activity and sedentary behaviour measures in the COMPASS study [Internet]. [COMPASS Technical Report Series.] Waterloo (ON): University of Waterloo; 2014 [cited 2021 Aug 18]. Available from: https://uwaterloo.ca/compass-system/publications/reliability-and-validity-physical-activity-and-sedentary

32. Zhuchkova S, Rotmistrov A. How to choose an approach to handling missing categorical data: (un)expected findings from a simulated statistical experiment. Qual Quant. 2022;56:1–22. https://doi.org/10.1007/s11135-021-01114-w

33. Ene M, Leighton EA, Blue GL, Bell BA. Multilevel Paper 3430-2015: Models for categorical data using SAS® PROC GLIMMIX: the basics. Paper presented at: SAS Global Forum 2015; 2015 Apr 26-29; Dallas, TX.

34. Rose G. Rose's strategy of preventive medicine. New York (NY): Oxford University Press Inc; 1992.

35. Audrain-McGovern J, Rodriguez D, Moss HB. Smoking progression and physical activity. Cancer Epidemiol Biomarkers Prev. 2003;12(11):1121-9.

36. Rodriguez D, Dunton GF, Tscherne J, Sass J. Physical activity and adolescent smoking: a moderated mediation model. Ment Health Phys Act. 2008; 1(1):17-25. https://doi.org/10.1016/j.mhpa.2008.04.001

37. Jachyra P. Boys, bodies, and bullying in health and physical education class: implications for participation and well-being. Asia-Pacific J Health Sport Phys Educ. 2016;7(2):121-38. https://doi.org/10.1080/18377122.2016.1196112

38. Frydendal S, Thing LF. A shameful affair? A figurational study of the change room and showering culture connected to physical education in Danish upper secondary schools. Sport Educ Soc. 2020;25(2):161-72. https://doi.org/10.1080/13573322.2018.1564654

39. Sandercock GR, Ogunleye A, Voss C. Associations between showering behaviours following physical education, physical activity and fitness in English schoolchildren. Eur J Sport Sci. 2016;16(1):128-34. https://doi.org/10.1080/17461391.2014.987321

40. Kirby J, Levin KA, Inchley J. Associations between the school environment and adolescent girls' physical activity. Health Educ Res. 2012;27(1): 101-14. https://doi.org/10.1093/her/cyr090

41. Dwyer JJ, Allison KR, Goldenberg ER, Fein AJ, Yoshida KK, Boutilier MA. Adolescent girls' perceived barriers to participation in physical activity. Adolescence. 2006;41(161):75-89.

42. Sperrin M, Martin GP. Multiple imputation with missing indicators as proxies for unmeasured variables: simulation study. BMC Med Res Methodol. 2020;20:185. https://doi.org/10.1186/s12874-020-01068-x

Other PHAC publications

Researchers from the Public Health Agency of Canada also contribute to work published in other journals. Look for the following articles published in 2022:

Brenner DR, Poirier A, Woods RR, […] **Demers AA**, et al. Projected estimates of cancer in Canada in 2022. CMAJ. 2022;194(17): E601-E607. https://doi.org/10.1503/CMAJ.212097

Liu L, **Pollock NJ**, **Contreras G**, **Tonmyr L**, **Thompson W**. Prevalence of suicidal ideation among adults in Canada: results of the second Survey on COVID-19 and Mental Health. Health Rep. 2022;33(5):13-21. https://doi.org/10.25318/82-003-x202200500002-eng

McGrath R, **Lang JJ**, Ortega FB, et al. Handgrip strength asymmetry is associated with slow gait speed and poorer standing balance in older Americans. Arch Gerontol Geriatr. 2022;102:104716. https://doi.org/10.1016/j.archger.2022.104716