

## Use of administrative record linkage to measure medical and social risk factors for early developmental vulnerability in Ontario, Canada

Natasha Ruth Saunders<sup>1-6,\*</sup>, Magdalena Janus<sup>6,7</sup>, Joan Porter<sup>3</sup>, Hong Lu<sup>3</sup>, Ashley Gaskin<sup>7</sup>, Gangamma Kalappa<sup>3</sup>, and Astrid Guttmann<sup>1-6</sup>

### Submission History

Submitted:	23/09/2020
Accepted:	04/12/2020
Published:	11/02/2021

<sup>1</sup>The Hospital for Sick Children, Toronto, Canada, M5G 1X8

<sup>2</sup>Department of Pediatrics, University of Toronto, Toronto, Canada, M5G 1X8

<sup>3</sup>ICES, Toronto, Canada, M4N 3M5

<sup>4</sup>Child Health Evaluative Sciences, SickKids Research Institute, Toronto, Canada, M5G 1X8

<sup>5</sup>Institute of Health Policy, Management and Evaluation, The University of Toronto, Toronto, Canada, M5T 3M6

<sup>6</sup>Edwin S. H. Leong Centre for Healthy Children, University of Toronto, Toronto, Canada, M5G 1X8

<sup>7</sup>Offord Centre for Child Studies, Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Canada, L8S 4K1

### Abstract

#### Background

Linkage of demographic, health, and developmental administrative data can enrich population-based surveillance and research on developmental and educational outcomes. Transparency of the record linkage process and results are required to assess potential biases.

#### Objectives

To describe the approach used to link records of kindergarten children from the Early Development Instrument (EDI) in Ontario to health administrative data and test differences in characteristics of children by linkage status. We demonstrate how socio-demographic and medical risk factors amass in their contribution to early developmental vulnerability and test the concordance of health diagnoses in both the EDI and health datasets of linked records.

#### Methods

Children with records in the 2015 EDI cycle were deterministically linked to a population registry in Ontario, Canada. We compared sociodemographic and developmental vulnerability data between linked and unlinked records. Among linked records, we examined the contribution of medical and social risk factors obtained from health administrative data to developmental vulnerability identified in the EDI using descriptive analyses.

#### Results

Of 135,937 EDI records, 106,217 (78.1%) linked deterministically to a child in the Ontario health registry using birth date, sex, and postal code. The linked cohort was representative of children who completed the EDI in age, sex, rural residence, immigrant status, language, and special needs status. Linked data underestimated children living in the lowest neighbourhood income quintile (standardized difference [SD] 0.10) and with higher vulnerability in physical health and well-being (SD 0.11), social competence (SD 0.10), and language and cognitive development (SD 0.12). Analysis of linked records showed developmental vulnerability is sometimes greater in children with social risk factors compared to those with medical risk factors. Common childhood conditions with records in health data were infrequently recorded in EDI records.

#### Conclusions

Linkage of early developmental and health administrative data, in the absence of a single unique identifier, can be successful with few systematic biases introduced. Cross-sectoral linkages can highlight the relative contribution of medical and social risk factors to developmental vulnerability and poor school achievement.

#### Keywords

Early Development Instrument; data linkage; vulnerability; social risk; child development

\*Corresponding Author:

Email Address: [Natasha.saunders@sickkids.ca](mailto:Natasha.saunders@sickkids.ca) (Natasha Ruth Saunders)

## Introduction

Routinely collected health and administrative data contain a wealth of information that can be used to inform health system delivery [1, 2]. Adding cross-sectoral data linkages, including datasets that capture developmental and social services records can enrich the contextual understanding of each of these sectors' data. Further, the relatively low costs of data linkage in comparison to traditional methods (e.g. survey collection) make linkages of multiple datasets appealing [3]. For young children specifically, there is a paucity of population level indicators of their development, health, well-being, and social risk beyond the neonatal period making linkages to a broad range of cross-sectoral datasets even more desirable for health and education system learnings [4].

In Canada, comprehensive health data are routinely collected in each province with every health system interaction for all legal residents with a valid universal provincial health insurance number. In Ontario, Canada's most populous province (approximately 14 million persons), these data are held securely at ICES (formerly the Institute for Clinical Evaluative Sciences), one of Canada's largest health data repositories. At ICES, unique encoded health identification numbers allow for linkage across health datasets and privacy legislation allows for their use for research. While use of secondary data for research has become increasingly attractive for researchers, it continues to have important methodological challenges [3]. With dataset linkage, in the absence of a unique identifier, errors are inevitable. Systematic errors in linkage, in turn, can lead to bias and threaten the reliability of results. Moreover, some records cannot be linked for a variety of reasons, including the incompleteness of the data used for linkage and the absence of identifiers common to both datasets [2, 3]. Study designs and interpretation can be strengthened with a good understanding of the linkage process, potential errors, and sources of bias. Journals increasingly encourage transparency in reporting studies using linked administrative data [1, 5]. While reporting linkage quality is an important component for quality assurance, linkage statistics are often not available or reported.

Despite the scarcity of population-level child health measures, one dataset that holds promise to enhance our understanding of child developmental health is the Early Development Instrument (EDI) [6]. The EDI is a valid and reliable teacher completed checklist which measures the developmental status of children in kindergarten. The EDI measures children's ability to meet age-appropriate developmental expectations in five domains: physical health and well being, social competence, emotional maturity, language and cognitive development, and communications skills and general knowledge [6]. The EDI, recognized and used internationally, focuses on the overall developmental outcomes for children as a health-relevant, measurable concept that has long-term consequences for individuals and populations [7]. Full details about the EDI, including its validation have been published elsewhere [8–16]. Linkage of these developmental data to health administrative data may be used to answer critical questions about the interplay between pre- and post-natal maternal and child health and social well-being and early childhood development [17–19].

Capitalizing on the rich child development data as measured by the EDI and the feasibility and importance of linkage with health and administrative data holdings available at ICES, this study aims to describe the record linkage strategy used to link one cycle of EDI with the health administrative data at ICES, estimate linkage rates, compare sociodemographic and health characteristics of the linked and unlinked data, and test the concordance of diagnoses in both the EDI and health datasets of those linked records. In addition, in the linked population, we describe the medical and social risk factor profile associated with EDI vulnerability outcomes.

## Methods

### Study design and setting

This was a population-based linkage study of kindergarten-aged children (ages 5 to 6 years) with a completed EDI record in 2015 using health data routinely collected from the publicly funded universal healthcare system and other administrative data available in Ontario, Canada. The datasets linked to the EDI used unique encoded identifiers and were analyzed at ICES, an independent, non-profit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyze health care and demographic data from designated health information custodians, without individual consent, for health system evaluation and improvement.

### The Early Development Instrument

The EDI is a 103-item validated questionnaire completed by senior kindergarten teachers in publicly funded schools in the second half of the school year that measures children's ability to meet age-appropriate developmental expectations across five domains: 1) physical health and well-being; 2) social competence; 3) emotional maturity; 4) language and cognitive development; and 5) communication skills and general knowledge [6, 20]. Designed as a population-level research tool, the EDI is a measure of developmental status of children across schools, neighbourhoods, as well as larger geographies, which with repeated implementations, provides indicators of developmental change over time in populations of children. It can be used to help identify cohorts of children who may be struggling, though not yet identified, and may benefit from universal programs. The 2015 implementation of the EDI in Ontario is the 4<sup>th</sup> provincial level collection in a cycle that began in 2004. The first cycle of EDI data constitutes the Ontario baseline for subsequent years and is used to determine the 10<sup>th</sup> percentile cut-offs for assessment of vulnerability. Children scoring below this 10<sup>th</sup> percentile cut-off are at higher odds of not meeting optimal developmental standards [6, 21, 22]. Such children may benefit from cost-effective, universal preventive programs [23]. The EDI also contains parent-reported and teacher-recorded medical and developmental diagnoses for children, including but not limited to, autism spectrum disorder (ASD), asthma, anxiety, cerebral palsy, and attention deficit and hyperactivity disorder (ADHD). These are selected from a list of over 30 diagnoses available to the teacher [23].

## Data sources

The Registered Person's Database (RPDB), the file of all residents in Ontario who are eligible for universal provincial health care coverage, was used to identify date of birth, sex and alpha-numeric residential postal code via health insurance information. Such coverage includes refugees and immigrants to Canada who have been granted permanent resident status. For health data, we used the Ontario Health Insurance Plan for outpatient physician billings, the National Ambulatory Care Reporting System for emergency department visits, and the Canadian Institute for Health Information's Discharge Abstract Database for inpatient hospitalizations. The MOMBABY database, an ICES-derived dataset, linked children born in Ontario hospitals to their mothers using birth hospitalization records [24]. We used the mother's encoded identifier to obtain pre- and post-natal information related to socio-demographic and medical factors of mothers. For children without records in the MOMBABY database, we linked to the Immigration, Refugee and Citizenship Canada's Permanent Resident Database [25] to determine whether the child entered Canada as an immigrant (e.g., economic class, family class, etc.) or refugee. Children who were not born in a hospital in Ontario or identified as an immigrant or refugee, were categorized as 'other', which includes children coming to Ontario through inter-provincial or -territorial migration and children born at home. We derived neighbourhood income quintile, the Ontario Marginalization Index material deprivation quintile [26], and rural residence by linking residential postal code with census data. The Ontario Drug Benefit database identified children and mothers who had ever used social assistance through filling a prescription paid for by the Ontario Drug Benefit Program [27].

## Study population

In Ontario, children enter Year 1 of kindergarten in September of the calendar year they turn 4 years old (junior kindergarten), and Year 2 the year they turn 5 (senior kindergarten). The majority of children (almost 90%) attend publicly funded schools [28]. All children in publicly funded Year 2 kindergarten ( $n = 135,937$ ) [29] in Ontario whose teacher completed the EDI questionnaire in the final term of the year (1<sup>st</sup> January to 30<sup>th</sup> June, 2015, age 5- to 6-years-old) were considered for linkage. Individuals were excluded if their EDI data did not have a minimum number of questions completed (i.e. valid for inclusion filter created by EDI developers that flags useable records) [23]. After linkage of the EDI to health and administrative data, we further excluded records that linked to the MOMBABY database but with a flag indicating a problem with linkage. For example, we excluded children where the infant's identifier was equivalent to the mother's identifier.

## Linkage methodology

We retrieved birth date and sex variables for all individuals with a valid Ontario health card number with a year of birth from 2006 to 2010 from the RPDB. These birth years were considered as potentially linkable with children who completed the 2015 EDI in Ontario. For all individuals in the retrieved cohort, we determined their postal code of residence as of

1<sup>st</sup> July, 2015 using the RPDB postal code file. We used deterministic linkage and variables of birth date, sex, and postal code to link EDI records to the RPDB file. For records that remained unlinked after this first pass, we expanded the catchment to include the postal code file for 2014. We performed this second pass to account for the possibility of an address change that might not have been reflected in the EDI or 2015 RPDB file. As in the first pass, birth date and sex variables for all individuals with a valid Ontario health card number with year of birth from 2006 to 2010 were retrieved from the RPDB file. For all individuals in the retrieved cohort, we used their postal code of residence as of 1<sup>st</sup> July, 2014 using the RPDB postal code file and deterministically linked them to the EDI records. Due to the lack of more specific identifiers in the EDI dataset, such as the child's name, we were unable to link records with identical values for birth date, sex, and postal code to the RPDB. Thus, multiple birth children of the same sex were excluded from the linked cohort. This approach was used as no unique identifier was available in each database that were common to both datasets.

## Outcomes

The outcome of interest was a child's vulnerability status as determined by their score on each of the five domains measured by the EDI. Children were considered to demonstrate overall vulnerability if they scored less than or equal to the 10<sup>th</sup> percentile cut-off for Ontario children on at least one domain.

## Covariates

We chose social and medical risk factors for a child's vulnerability based on previous research that has shown a relationship between these factors [30–35]. Risk factors were categorized as medical or social (Supplementary Appendix A). Medical risk factors related to the birth event (premature birth, low birthweight, neonatal intensive care unit stay at birth) could only be determined for children born in an Ontario hospital. Because extreme prematurity (<28 weeks gestation) and extremely low birth weight (<1500 grams) can contribute to significant neurodevelopmental morbidity, these risk factors were given double the weight of children born preterm at 28 to ≤36 weeks gestation or low birth weight at 1500 to <2500 grams, and counted as two risk factors each. Other medical risk factors included congenital anomalies, major illnesses, hospitalizations for 6 or more days, and any intensive care unit stay. To determine the presence of a major illness, we used The Johns Hopkins ACG System Version 10 [36, 37] a case-mix methodology for aggregating conditions into adjusted diagnostic groups (ADGs) that are similar in expected resource use to determine whether children had 2 or more major diagnoses from the first of their birth or their first eligibility for provincial health insurance to their 4<sup>th</sup> birthday.

Sociodemographic risk factors were derived from both the child's and, for those born in Ontario, their mother's encounters with the health system, and involved linkages with census, physician billing, hospitalization, immigration, Ontario Drug Benefit program, and the Narcotic Monitoring System sources [25, 38]. These risk factors include: child or mother was a refugee; child was discharged to social services at birth; child had an outpatient visit or hospitalization for maltreatment;

child lives in the lowest neighbourhood income quintile; mother was younger than 19 years of age at first birth; mother had 4 or more children as of the child's 4<sup>th</sup> birthday; mother has 10 or more ADGs in the post-natal period; mother has a history of mental illness in the pre- or post-natal period; and mother has a history of an emergency department visit or hospitalization as a result of experiencing violence.

For each child we calculated a score representing the sum of risk factors in each of the medical and social domains where, for example, a score of 0 under the social domain meant the child had no social risk factors. We categorized scores in each of the domains (medical or social) as zero, one, two, and three or more and then created 16 dyads representing a child's medical and sociodemographic risk factor experience. A dyad of '00' represented a child with no medical or sociodemographic risk factors while the dyad '33' represented a child with both medical and social risk scores of three or more.

## Concordance of education and health data for common clinical conditions

We selected five common clinical conditions reported in the EDI by the teacher, often from information provided by the parent, to assess concordance between the EDI-based reports and health conditions recorded in administrative data. Conditions included anxiety, asthma, ADHD, ASD, and cerebral palsy (CP). To determine whether a child had the condition according to administrative data, we observed from when the child had provincial health insurance (birth or migration to Ontario) to 2 years following the completion of the EDI and where the diagnosis was reported in either physician billings, emergency department and hospital discharge records, or in the narcotic monitoring records (ADHD only). For asthma, we used an asthma specific database developed at ICES and compiled based on a validated algorithm for identifying this condition in administrative data [39]. Data sources and definitions of medical diagnoses are provided in Supplementary Appendix A.

## Analysis

### Linkage analysis

To assess potential biases and limitations of the EDI data, we compared the linked EDI data to the unlinked EDI data in a descriptive analysis of variables recorded in the EDI, as well as information obtained from postal code linkages. We used the standardized difference of the mean to test for statistically significant differences where a value equal to or greater than 0.10 was considered significant.

### Vulnerability analysis

We used descriptive analysis to better understand the relationship between medical and social risk factors and a child's vulnerability. We plotted risk factor dyads against vulnerability, defined as scoring at or below the 10<sup>th</sup> percentile on at least one of the five domains or scoring at or below the 10<sup>th</sup> percentile on two or more domains.

We examined the concordance of health conditions reported on the EDI with the administrative data identification of five health conditions. We calculated the agreement (kappa) between EDI and administrative data in identifying the condition, as well as the positive predictive value of the EDI source to identify a condition relative to the administrative data reference standard.

All data analyses were conducted at ICES using SAS Enterprise Guide 7.1 (SAS Institute Inc., Cary, North Carolina).

This study was approved by the Research Ethics Board at The Hospital for Sick Children, Toronto, Ontario.

## Results

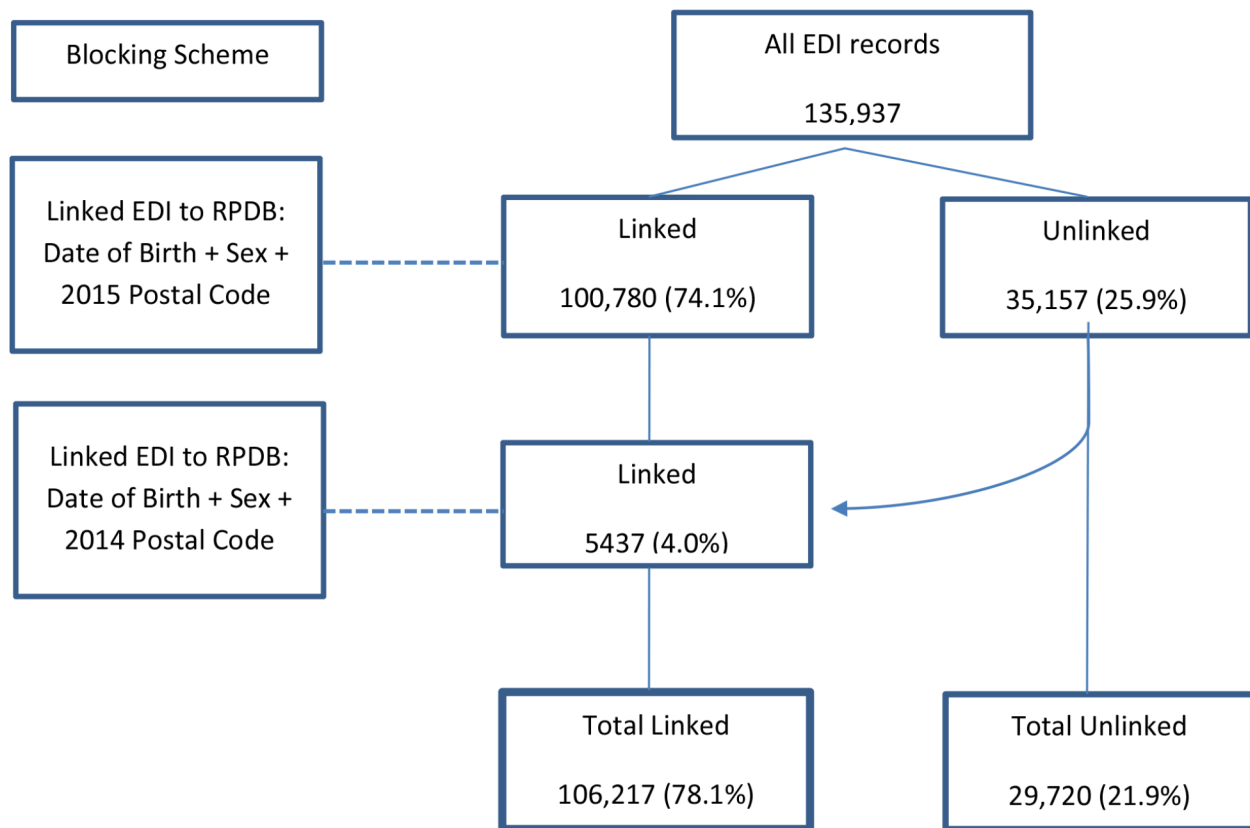
There were 135,937 EDI records completed in Ontario in 2015 available for linkage. Following deterministic linkage with birth date, sex and postal code, 100,780 (74.1%) of files were linked after the first pass using the 2015 RPDB postal code file and a further 5,437 (4.0%) files were linked after the second pass using the 2014 RPDB postal code file. This culminated in a total of 106,217 (78.1%) files being linked (Figure 1). A further 2,062 files were excluded due to invalid EDI records and 207 files were excluded due to invalid identifiers when linked to maternal record and birth hospitalization for those born in Ontario hospitals (Figure 2). Through this linkage process, 4.2% of EDI records represented multiple birth children of the same sex and thus could not be linked due to the lack of more specific identifiers in the EDI dataset.

### Linked versus unlinked records

Characteristics of children comparing those with linked and unlinked records and only those with linked records included in the study are shown in Tables 1 and 2, respectively. Supplementary Appendix B further details baseline characteristics of linked children who completed the EDI and were included in the study. Children with linked and unlinked records had similar sociodemographic characteristics including age, sex, and rural residence (Table 1). Significant differences were observed between groups by neighbourhood level income quintile with lower linkage rates for children living in the lowest neighbourhood level income quintile (income quintile 1: 22.5% unlinked vs. 18.3% linked; standardized difference [SD] 0.10) and higher linkage rates for those who had been at the school administering the EDI for longer than one month (98.9% linked vs. 94.4% unlinked; SD 0.25). There were no differences between children with linked and unlinked records by special needs designation, mother tongue, by enrollment in French Immersion education, or by positive multiple challenge index assignment (defined as scoring below standardized cut-point on 9 or more of the 16 subdomains). Differences in mean domain scores were observed based on linkage status in physical health and well-being ( $8.76 \pm 1.40$  linked,  $8.61 \pm 1.49$  unlinked; standardized difference 0.11), social competence ( $8.21 \pm 1.97$  linked,  $8.01 \pm 2.06$  unlinked; SD 0.10), and language and cognitive development ( $8.82 \pm 1.69$  linked,  $8.61 \pm 1.82$  unlinked; SD 0.12) domains, with scores being higher in the children with linked records. Differences were not observed between groups in emotional maturity



Figure 1: Deterministic linkage process used to link the 2015 Early Development Instrument (EDI) to the Ontario Registered Persons Database (RPDB)



and communication skills and general knowledge domains. When comparing linkage groups based on the proportion below the 10<sup>th</sup> percentile in any domain, no differences were observed.

### Concordance of health administrative data and EDI for select child health diagnoses

Table 3 shows the concordance between select parent reported/teacher recorded medical diagnoses in the EDI and health administrative data records of those same diagnoses. Generally, anxiety, asthma, cerebral palsy, and ADHD were infrequently recorded in the EDI when there was a record in health administrative data. For ASD, records in the EDI occurred in almost half of health administrative data records. ASD showed good concordance with a positive predictive value of 89.4% and a negative predictive value of 98.4% and kappa 0.60. There was generally fair concordance with EDI and administrative data in identifying cerebral palsy with positive and negative predictive values of 91.8% and 99.7% respectively and kappa 0.34.

### Medical risks

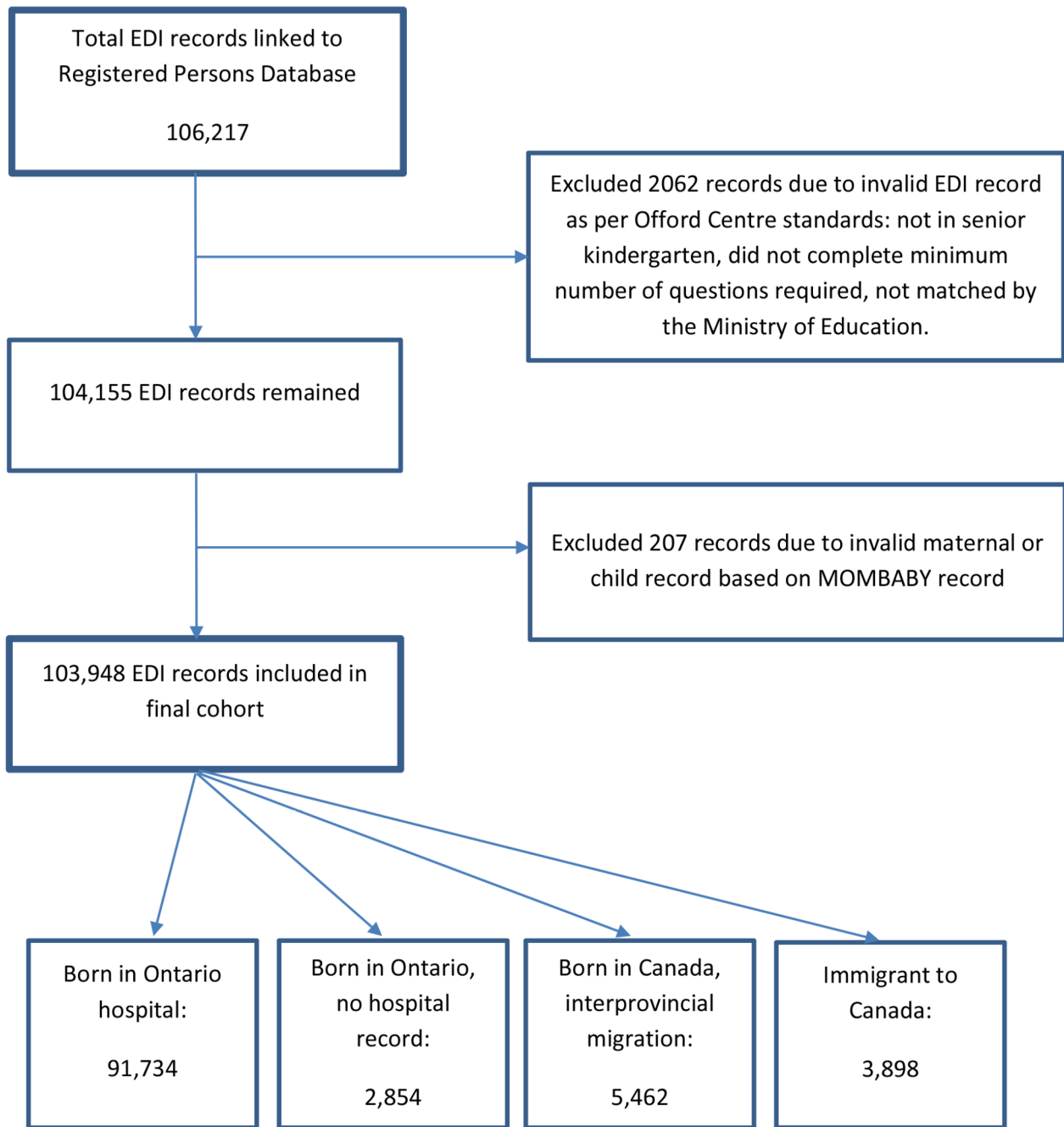
Among linked individuals, 31,530 (30.3%) demonstrated overall vulnerability (one or more domains), among whom 15,039 (14.4%), 7,240 (7.0%), 4,392 (4.2%), 2,852 (2.7%), and 2,007 (1.9%) had vulnerabilities on 1, 2, 3, 4, and 5 domains, respectively (Table 4). For children born very

preterm (< 28 weeks gestation) and preterm (28 to 36 weeks), 54.1% and 36.6%, respectively, were vulnerable on one or more domains; with similar proportions observed in children born with very low or low birth weight. Among children with a prolonged neonatal intensive care unit stay at birth, congenital anomalies, major childhood illnesses, prolonged hospitalizations, and any intensive care unit stay, between 40.1% to 46.0% demonstrated overall vulnerability.

### Social risks

There were 3,366 children who were either themselves or had mothers who were refugees, of whom 1,247 (37.0%) were vulnerable on one or more domains. Out of only 78 children who were discharged to social services at birth, 67.9% were vulnerable on one or more domains (Table 4). Among children with a history of maltreatment, 45.6% were vulnerable on at least one domain. Across all neighbourhood income quintiles, there was a gradient with increasing number of vulnerable domains with decreasing neighbourhood income quintile, with the highest percentage of children (40.0%) with overall vulnerability in the lowest income quintile neighbourhoods. Among children with family members (either themselves or their mother) who received income assistance, 53.9% were vulnerable on at least one domain. About a third of children with mothers with multiple (10+) medical comorbidities (33.7%) and mental illness (33.1%) were overall vulnerable (Table 4). Among only 63 children in the sample with known

Figure 2: Flow chart of inclusion and exclusions of Early Development Instrument (EDI) records for assessing social and medical risk factors for vulnerability



history of maternal assault, 65.1% of children demonstrated vulnerability.

### Medical and social risk

Supplementary Appendix C shows the number and proportion of the population affected by medical and social risk factors. Most (85.3%) children had no medical risk factors with only 5.3% having two or more medical risk factors. In contrast, 35.1% of children had no social risk factors and 34.1% had two or more social risk factors. Figure 3 shows the proportion of individuals with vulnerability on one or more EDI domains by number of social and medical risk factors. With increasing social or medical risk factors, there

was an increase in the proportion with EDI vulnerabilities. Among those with three social risk factors, regardless of the number of medical risk factors, the proportion with overall vulnerability (47.5%–61.7%) was higher than all other combined social and medical risk groups. Among those with three social risk factors and no medical risk factors, 47.5% were vulnerable overall compared with 35.5% of those with three or more medical risk factors with no social risk factors.

Table 1: Characteristics of children with linked and unlinked early development instrument records

	Linked	Unlinked	Standardized difference	Linkage rate (%)
Characteristic	N=106,217	N=29,720		
Sex, male, n (%)	54,810 (51.6)	14,916 (50.2)	0.03	78.6
Age, mean ( $\pm$ SD)	5.68 $\pm$ 0.29	5.68 $\pm$ 0.30	0.02	
Rural residence, yes, n (%)	11,015 (10.4)	2,570 (8.6)	0.06	81.1
Neighbourhood Income quintile n (%)				
1 (lowest)	19,476 (18.3)	6,694 (22.5)	0.10	74.4
2	19,307 (18.2)	5,826 (19.6)	0.04	76.8
3	21,277 (20.0)	5,777 (19.4)	0.01	78.6
4	25,120 (23.6)	6,098 (20.5)	0.08	80.5
5 (highest)	20,824 (19.6)	5,106 (17.2)	0.06	80.3
Missing	213 (0.2)	88 (0.3)	0.02	70.8
Child in class for >1 month, n (%)	105,049 (98.9)	28,053(94.4)	0.25	78.9
Special Needs, yes, n (%)	4,531 (4.3)	1,155 (3.9)	0.02	79.7
Child's first language, n (%)				
English	83,373 (78.5)	22,889 (77.0)	0.04	78.5
French	3,217 (3.0)	1,025 (3.4)	0.02	75.8
Other	19,627 (18.5)	5,806 (19.5)	0.03	77.2
English or French as a second language, n (%)	9,806 (9.2)	2,928 (9.9)	0.02	77.0
English as second language program	2,282 (2.1)	825 (2.8)	0.04	73.4
French as second language program				
French Immersion Program, yes, n (%)	13,265 (12.5)	3,219 (10.8)	0.05	80.5
Multiple Challenge Index, yes, n (%)	4,944 (4.7)	1,553 (5.2)	0.03	76.1
EDI Domain scores, mean ( $\pm$ SD)				
Physical health and well-being	8.76 $\pm$ 1.40	8.61 $\pm$ 1.49	0.11	
Social competence	8.21 $\pm$ 1.97	8.01 $\pm$ 2.06	0.10	
Emotional health and maturity	7.95 $\pm$ 1.65	7.80 $\pm$ 1.71	0.09	
Language and cognitive development	8.82 $\pm$ 1.69	8.61 $\pm$ 1.82	0.12	
Communication skills and general knowledge	7.83 $\pm$ 2.56	7.64 $\pm$ 2.64	0.07	
Below 10 <sup>th</sup> percentile on at least 1 domain score, n (%)	31,938 (30.1)	10,120 (34.1)	0.09	75.9
Below 10 <sup>th</sup> percentile on 2+ domain scores, n (%)	16,730 (15.8)	5,406 (18.2)	0.07	75.6
Vulnerable, Below 10 <sup>th</sup> percentile in domain, n (%)				
Physical health and well-being	17,791 (16.7)	5,828 (19.6)	0.07	75.3
Social competence	12,574 (11.8)	4,018 (13.5)	0.05	75.8
Emotional health and maturity	14,233 (13.4)	4,419 (14.9)	0.04	76.3
Language and cognitive development	8,003 (7.5)	2,737 (9.2)	0.06	74.5
Communication skills and general knowledge	12,457 (11.7)	3,822 (12.9)	0.03	76.5
Special concerns, yes, n (%)				
Physical disability	1,336 (1.3)	350 (1.2)	0.01	79.2
Visual impairment	911 (0.9)	272 (0.9)	0.01	77.0
Hearing impairment	902 (0.8)	239 (0.8)	0.00	79.1
Speech impairment	6,326 (6.0)	1,652 (5.6)	0.02	79.3
Learning disability	3,793 (3.6)	1,107 (3.7)	0.01	77.4
Emotional problem	3,991 (3.8)	1,340 (4.5)	0.04	74.9
Behavioural problem	5,820 (5.5)	1,883 (6.3)	0.04	75.6
Home environment	2,901 (2.7)	1,300 (4.4)	0.09	69.1
Chronic medical problems	1,350 (1.3)	349 (1.2)	0.01	79.5
Dental problems	443 (0.4)	174 (0.6)	0.02	71.8
Other problem	3,233 (3.0)	920 (3.1)	0.00	77.8
Attention deficit hyperactivity disorder, yes, n (%)	553 (0.5)	176 (0.6)	0.01	75.9
Autism spectrum disorder, yes, n (%)	1,527 (1.4)	345 (1.2)	0.02	81.6

## Discussion

In this large population-level linkage study, we described successful linkage of school-based developmental with health

administrative databases for kindergarten-aged children in Ontario, Canada and demonstrated how we can measure the relationship between medical and social risk factors and

Table 2: Characteristics of the study cohort used to assess social and medical risk factors for vulnerability

<b>Socio-demographics</b>	
Sex	
Female, n (%)	50,308 (48.4%)
Male, n (%)	53,640 (51.6%)
Age at completion of EDI	
Mean $\pm$ SD	5.68 $\pm$ 0.29
Median (IQR)	6 (5-6)
Place of birth	
Ontario	91,734 (88.2)
Immigrant	3,898 (3.7)
Home Birth	2,854 (2.7)
Inter-provincial migrant	5,462 (5.2)
<b>Rural residence</b>	
No, n (%)	93,996 (90.4%)
Yes, n (%)	9,931 (9.6%)
Missing	21 (0.0%)
Ontario Marginalization Index	
Material Deprivation Quintile	
Quintile 1 (lowest), n (%)	24,199 (23.3%)
Quintile 2, n (%)	22,636 (21.8%)
Quintile 3, n (%)	19,314 (18.6%)
Quintile 4, n (%)	17,729 (17.1%)
Quintile 5, n (%)	19,768 (19.0%)
Missing, n (%)	302 (0.3%)
<b>Medical risk factors</b>	
Congenital anomalies	
No, n (%)	99,337 (95.6%)
Yes, n (%)	2,004 (1.9%)
Missing	2,607 (2.5%)
ADG minor illness child	
No, n (%)	92,944 (89.4%)
Yes, n (%)	8,397 (8.1%)
Missing	2,607 (2.5%)
ADG major illness child	
No, n (%)	94,728 (91.1%)
Yes, n (%)	6,613 (6.4%)
Missing	2,607 (2.5%)
Prolonged hospitalization (6+ days)	
No, n (%)	96,861 (93.2%)
Yes, n (%)	4,480 (4.3%)
Missing	2,607 (2.5%)
Any ICU hospitalization	
No, n (%)	99,216 (95.4%)
Yes, n (%)	2,125 (2.0%)
Missing	2,607 (2.5%)
<b>Social risk factors</b>	
History of maltreatment	
No	102,486 (98.6%)
Yes	1,462 (1.4%)
Income Assistance	
No, n (%)	90,732 (87.3%)
Yes, n (%)	13,216 (12.7%)

Continued



Table 2: Continued.

<b>Health care use</b>	
Annual physician visits	
Mean $\pm$ SD	6.20 $\pm$ 3.74
Median (IQR)	5 (4-8)
Received 18-month enhanced well baby visit	
No, n (%)	54,987 (52.9%)
Yes, n (%)	46,354 (44.6%)
Missing	2,607 (2.5%)

Table 3: Agreement of conditions commonly reported in the early development instrument with the condition's identification in administrative data

		<b>Administrative health data</b>			<b>Concordance</b>	
		<b>Yes</b>	<b>No</b>	<b>Total</b>		
		<b>Anxiety</b>				
<b>EDI</b>	<b>Yes</b>	60	63	123	PPV	48.8%
	<b>No</b>	8,618	95,207	103,825	NPV	91.6%
	<b>Total</b>	8,678	95,270	103,948	Kappa (95% CI)	0.01 (0.01 – 0.01)
		<b>Asthma</b>				
<b>EDI</b>	<b>Yes</b>	107	7	114	PPV	93.9%
	<b>No</b>	30,622	73,212	103,834	NPV	70.4%
	<b>Total</b>	30,729	73,219	103,948	Kappa (95% CI)	0.01 (0.00 – 0.01)
		<b>Attention Deficit Hyperactivity Disorder</b>				
<b>EDI</b>	<b>Yes</b>	340	209	549	PPV	61.9%
	<b>No</b>	3,442	99,957	103,399	NPV	96.7%
	<b>Total</b>	3,782	100,166	103,948	Kappa (95% CI)	0.15 (0.14 – 0.16)
		<b>Autism Spectrum Disorder</b>				
<b>EDI</b>	<b>Yes</b>	1,343	160	1,503	PPV	89.4%
	<b>No</b>	1,600	100,845	102,445	NPV	98.4%
	<b>Total</b>	2,943	101,005	103,948	Kappa (95% CI)	0.60 (0.58 – 0.61)
		<b>Cerebral Palsy</b>				
<b>EDI</b>	<b>Yes</b>	89	8	97	PPV	91.8%
	<b>No</b>	336	103,515	103,851	NPV	99.7%
	<b>Total</b>	425	103,523	103,948	Kappa (95% CI)	0.34 (0.29 – 0.39)

PPV – positive predictive value, NPV – negative predictive value, CI – confidence interval.

vulnerability. Overall linkage rates for the 2015 Ontario EDI and RPDB databases were good at 78.1% using deterministic linkage with exact match on birth date, sex, and postal code. Comparison of the linked and unlinked files suggests few systematic biases exist when linking these data. No differences between children with linked and unlinked records were observed for most socio-demographic characteristics and the proportion of children considered vulnerable in any domain. However, we showed linked data marginally under-represented the proportion of children in the lowest neighbourhood income quintile and compared to unlinked, children with linked records had slightly higher mean domain scores for physical health and well-being, social competence, and language and cognitive development. Comparison of teacher-reported data with administrative health data for specific disease cohorts suggests that universally, health conditions identified in administrative health data were infrequently recorded in the EDI but there

was high positive predictive value for some cohorts, including asthma, ASD, and cerebral palsy. Among children with medical risk, as well as among those with social risks, there were higher percentages of children with developmental vulnerabilities than among those without risks. In addition, our results showed that children with three or more social risk factors, regardless of the number of medical risk factors had a greater proportion who were vulnerable compared to children with any number of medical risk factors.

The overall linkage rate for the EDI and RPDB data was comparable to other record linkage studies using health and administrative data [25, 40, 41]. The linkage statistics are important for understanding potential biases in future studies using school-based data linked to health data and are increasingly required by researchers for publication [42]. Children living in the lowest income neighbourhoods and with greatest vulnerability scores had the lowest linkage rates.

Table 4: Number of domains with reported vulnerability for the 2015 early development instrument cohort overall and by medical and social risk factors

	Number of domains with vulnerability (<10th percentile), n (%)						Total
	0	1	2	3	4	5	
<b>Total</b>	72,418 (69.7)	15,039 (14.4)	7,240 (7.0)	4,392(4.2)	2,852 (2.7)	2,007 (1.9)	103,948 (100)
<b>Medical Risk Factors</b>							
*Prematurity							
Very preterm (< 28 weeks)	100 (45.9)	38 (17.4)	29 (13.3)	23 (10.6)	14 (6.4)	14 (6.4)	218 (100.0)
Preterm (28 to 36 weeks)	3,871 (63.4)	988 (16.2)	485 (7.9)	336 (5.5)	241 (3.9)	185 (3.0)	6,106 (100.0)
*Birth weight							
Very low birth weight (< 1500 g)	309 (51.0)	99 (16.3)	64 (10.6)	56 (9.2)	41 (6.8)	37 (6.1)	606 (100.0)
Low birth weight (< 2500 g)	2,544 (63.0)	639 (15.8)	325 (8.1)	231 (5.7)	161 (4.0)	135 (3.3)	4,035 (100.0)
*Neonatal ICU at birth (3+ days)	3,387 (59.9)	881 (15.6)	514 (9.1)	366 (6.5)	278 (4.9)	228 (4.0)	5,654 (100.0)
Congenital anomaly	1,083 (54.0)	311 (15.5)	180 (9.0)	144 (7.2)	140 (7.0)	146 (7.3)	2,004 (100.0)
Major illness	3,759 (56.8)	1,033 (15.6)	595 (9.0)	416 (6.3)	414 (6.3)	396 (6.0)	6,613 (100.0)
Prolonged hospitalization (6+ days)	2,562 (57.2)	748 (16.7)	382 (8.5)	285 (6.4)	249 (5.6)	254 (5.7)	4,480 (100.0)
Any ICU hospitalization	1,170 (55.1)	350 (16.5)	203 (9.6)	136 (6.4)	136 (6.4)	130 (6.1)	2,125 (100.0)
<b>Social Risk Factors</b>							
Child or mother is a refugee	2,119 (63.0)	527 (15.7)	308 (9.2)	196 (5.8)	121 (3.6)	95 (2.8)	3,366 (100.0)
*Infant discharged to social services at birth	25 (32.1)	14 (17.9)	16 (20.5)	8 (10.3)	6 (7.7)	9 (11.5)	78 (100.0)
Child history of maltreatment	796 (54.4)	267 (18.3)	153 (10.5)	113 (7.7)	72 (4.9)	61 (4.2)	1,462 (100.0)
Low neighbourhood level income							
Quintile 1 (lowest)	11,334 (60.0)	3,228 (17.1)	1,772 (9.4)	1,150 (6.1)	791 (4.2)	610 (3.2)	18,885 (100.0)
Quintile 2	12,458 (66.0)	3,047 (16.1)	1,459 (7.7)	876 (4.6)	610 (3.2)	435 (2.3)	18,885 (100.0)
Quintile 3	15,051 (71.0)	2,920 (13.8)	1,433 (6.8)	888 (4.2)	549 (2.6)	364 (1.7)	21,205 (100.0)
Quintile 4	16,837 (73.2)	3,092 (13.4)	1,368 (5.9)	845 (3.7)	515 (2.2)	358 (1.6)	23,015 (100.0)
Quintile 5 (highest)	16,695 (76.2)	2,748 (12.6)	1,203 (5.5)	628 (2.9)	384 (1.8)	238 (1.1)	21,896 (100.0)
Income Assistance	6,232 (47.2)	2,637 (20.0)	1,597 (12.1)	1,202 (9.1)	831 (6.3)	717 (5.4)	13,216 (100.0)
*Mother's age at first birth < 19 years	2,227 (47.6)	989 (21.1)	579 (12.4)	392 (8.4)	270 (5.8)	220 (4.7)	4,677 (100.0)
*Mother with 4+ children	3,371 (58.8)	1,041 (18.2)	537 (9.4)	343 (6.0)	255 (4.4)	184 (3.2)	5,731 (100.0)
*Mother 10+ medical comorbidities	24,032 (66.3)	5,677 (15.7)	2,785 (7.7)	1,777 (4.9)	1,142 (3.2)	826 (2.3)	36,239 (100.0)
*Mother history of mental illness	28,255 (66.9)	6,491 (15.4)	3,149 (7.5)	2,025 (4.8)	1,327 (3.1)	1,004 (2.4)	42,251 (100.0)
*Mother history of assault	22 (34.9)	15 (23.8)	12 (19.0)	-	-	7 (11.1)	63 (100.0)

Small cell sizes (< 6) suppressed as per institutional policy.

\*Data only available for children born in Ontario as variable based on birth hospitalization record or maternal record linkage with infant birth hospitalization.

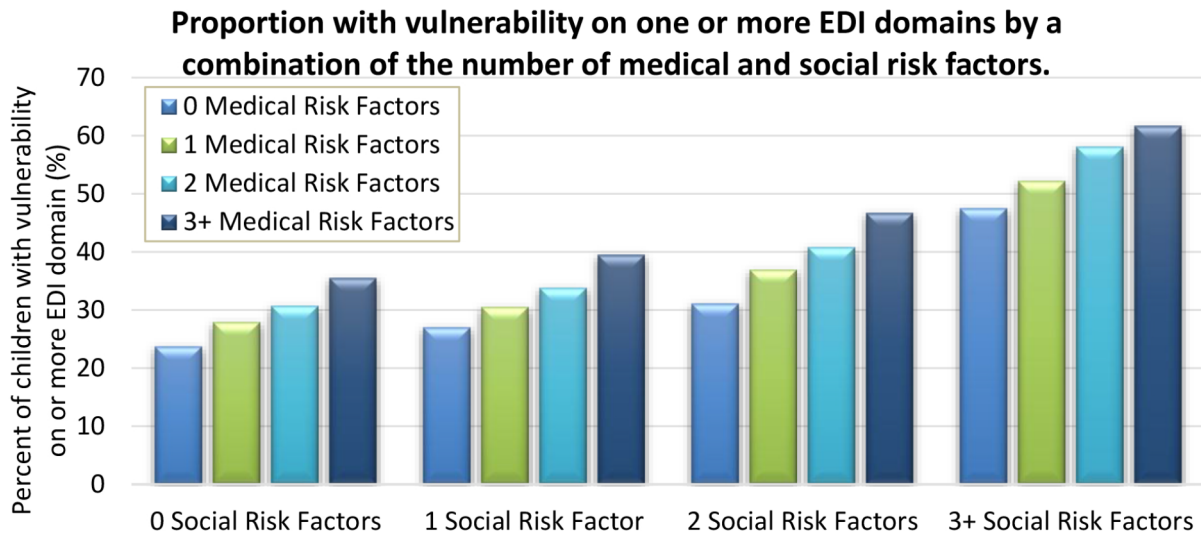
Others have found that incomplete records were less likely to match and were often related to social disadvantage [43, 44]. Frequent residential mobility, which is common in early childhood [45, 46], is associated with increased risk of poor physical and/or mental health and such children are overrepresented in low income families, children with no fully employed adult in the household, and children in single-parent households [45–47]. Residential mobility, including a short period of time in class (< one month) before the EDI completion, may have contributed to unlinked records and vulnerability findings based on our inability to match on postal code in the RPDB and EDI where updates to each database may not occur simultaneously. Adding the 2014 postal code file mitigated this with a 4% improvement in linkage rates. Further work to advance linkage using surname and first name identifiers would likely improve linkage but these are not currently available with existing government and data protection regulations for use of individual-level data.

We explored the concordance of records of medical conditions between EDI databases and administrative health records, acknowledging that neither can be considered definitive. The positive predictive values were lowest for anxiety (48.8%) and ADHD (61.9%). In contrast, ASD had a positive predictive value of 89.4%. This may be explained by

several factors. First, while some conditions are symptomatic in kindergarten-age children, they may not have been brought to the attention of a teacher yet and therefore not recorded in the EDI. Conversely, in some situations, conditions may be communicated to teachers before a medical diagnosis is made. With ADHD and anxiety, for example, symptoms are often present in young children but formal diagnosis and reporting to the school may not occur until primary grades [48]. With asthma, though highly prevalent in young children, because it does not typically affect day-to-day learning, a teacher may not be made aware of a child's condition, especially if mild [49]. In contrast, ASD is often diagnosed by age 4 years [50], prior to Year 2 (senior) kindergarten, and typically requires additional developmental supports of which the teacher is aware. Thus, it may be more commonly recorded in the EDI relative to other childhood conditions.

In children with at least one medical risk factor and no social risk factors, 27.9% were vulnerable on at least one domain, a proportion that increased to 35.5% when three or more medical risk factors were present. These are proportions similar to those reported in studies in Manitoba, Canada and South Australia of children with perinatal medical risk factors (e.g. low birth weight and neonatal intensive care unit stay) [30, 51]. For such children, robust and universal

Figure 3: Proportion of kindergarten students with linked Early Development Instrument records with vulnerability in one or more domains by the number of social and medical risk factors



systems exist to ensure an optimal developmental trajectory, particularly for those who have neonatal risk with established provincial or national neonatal neurodevelopmental follow up systems [52, 53]. In contrast, among children with at least one social risk factor, 27.0% were vulnerable on at least one domain, and where there were three social risk factors, this number rose to 47.5% - all in the absence of medical risk factors. While some universal programming to optimize developmental trajectories exist for those with social risk (e.g. Ontario's Health Babies Healthy Children, British Columbia's Early Years Program), they do not match the comprehensiveness and strength of established programs for medical risk in Ontario and elsewhere in Canada [54, 55]. Our findings showing the relative importance of cumulative social risks in comparison with medical ones in a population of kindergarten children in Ontario, together with studies from across North America and Australia, point to the universality of these phenomena in children's developmental trajectories [30, 51, 56, 57]. Further, we have shown the proportion of the population with social compared with medical risk factors is large. Such a finding underscores the importance of evaluation of social risk factors in clinical assessments to identify young children at risk of poor health and educational outcomes [58]. Using population level developmental and health data to understand how social and medical risks stack up to contribute to vulnerability highlights where gaps and opportunities exist to enhance and target programming for at-risk children to optimize early development.

While linkage of developmental and health data offers unprecedented opportunities to elucidate the impact of prenatal and postnatal social and medical exposures on early childhood development, our linkage is not without limitations. EDI data are collected in publicly funded schools only. Even though in Ontario this means inclusion of about 90% of the cohort, it does exclude data on children attending private schools or schools in First Nations communities [28]. There are also inherent limitations of administrative databases: diagnoses made and services delivered outside the medical

billing structure are not included, and children in unstable living circumstances (e.g. children discharged to social services from the birth hospitalization) may not have medical records and thus their records could not be linked thereby underestimating the full impact of social risk on vulnerability measured in the EDI. All these factors may have limited the generalizability of our findings, even though they are similar to findings with other cohorts. In addition, some social and medical risk factors are based on maternal or perinatal information. Consequently, we underestimated the true burden of such risk factors for immigrants, interprovincial migrants, and children born out of hospital. We developed weighted measures of medical risk, based on clinical plausibility and feasibility of obtaining such information using administrative data though such weights have not been validated.

## Conclusions

Deterministic linkage of school-based developmental and health administrative data allows for efficiency in augmenting our understanding of pathways towards healthy early childhood development. Good linkage rates suggest deterministic linkage is adequate to link EDI with health data with small systematic biases in linked versus unlinked records using very few identifiers. Knowledge of potential bias in unlinked records is becoming increasingly important and we have shown that linked EDI data provide a representative resource for understanding early child development using health and demographic data. As in other jurisdictions, we have shown social risk plays an important role in childhood development and should be a priority area for intervention.

## Acknowledgments

We thank Mahmoud Azimae for his assistance with linkage of EDI and RPDB datasets.

The Early Development Instrument (EDI) is a questionnaire developed by Dr. Dan Offord and Dr. Magdalena Janus at

the Offord Centre for Child Studies at McMaster University, Hamilton, Ontario, Canada. The EDI is a teacher-completed checklist that measures children's developmental health at school entry in kindergarten in five domains: physical health and well-being, social competence, emotional maturity, language and cognitive development and communication skills and general knowledge. We gratefully acknowledge all the teachers who committed their time and energy in completing the EDI questionnaires.

We thank IMS Brogan Inc. for use of their Drug Information Database.

## Statement on conflicts of interest

All authors have no conflicts of interest to declare.

## Financial support

This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by ICES or the Ontario MOHLTC is intended or should be inferred. Parts of this material are based on data and information compiled and provided by the Canadian Institute for Health Information (CIHI) and Immigration, Refugees and Citizenship Canada (IRCC). However, the analyses, conclusions, opinions and statements expressed herein are those of the authors, and not necessarily those of CIHI and IRCC.

## Data sharing

The data set from this study is held securely in coded form at ICES. Data-sharing agreements prohibit ICES from making the data set publicly available, but access may be granted to those who meet pre-specified criteria for confidential access, available at [www.ices.on.ca/DAS](http://www.ices.on.ca/DAS). The full data set creation plan and underlying analytic code are available from the authors upon request, understanding that the programs may rely upon coding templates or macros that are unique to ICES.

## Contributors statement

N. Saunders conceptualized and designed the study, interpreted the results, drafted the initial manuscript, revised the manuscript, and approved the final manuscript as submitted. J. Porter, M. Janus, A. Gaskin and A. Guttmann conceptualized and designed the study, interpreted the results, revised the manuscript, and approved the final manuscript as submitted. H. Lu and G. Kalappa had access to and analyzed the data, interpreted the results, revised the manuscript, and approved the final manuscript as submitted. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

## References

1. Nicholls SG, Langan SM, Benchimol EI. Routinely collected data: the importance of high-quality diagnostic coding to research. *CMAJ*. 2017;189(33):E1054–E5. <https://doi.org/10.1503/cmaj.170807>
2. Ford E, Boyd A, Bowles JKF, et al. Our data, our society, our health: A vision for inclusive and transparent health data science in the United Kingdom and beyond. *Learn Health Syst*. 2019;3(3):e10191. <https://doi.org/10.1002/lrh2.10191>
3. Harron K, Dibben C, Boyd J, et al. Challenges in administrative data linkage for research. *Big Data Soc*. 2017;4(2):2053951717745678. <https://doi.org/10.1177/2053951717745678>
4. Clark H, Coll-Seck AM, Banerjee A, et al. A future for the world's children? A WHO-UNICEF-Lancet Commission. *Lancet*. 2020;395(10224):605–58. [https://doi.org/10.1016/S0140-6736\(19\)32540-1](https://doi.org/10.1016/S0140-6736(19)32540-1)
5. Benchimol EI, Langan S, Guttmann A, et al. Call to RECORD: the need for complete reporting of research using routinely collected health data. *J Clin Epidemiol*. 2013;66(7):703–5. <https://doi.org/10.1016/j.jclinepi.2012.09.006>
6. Janus M, Offord DR. Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*. 2007; 39(1):1–22. <https://doi.org/10.1037/cjbs2007001>
7. Janus M, Reid-Westoby C. Monitoring the development of all children: the Early Development Instrument. *Early Childhood Matters*. 2016;7:40–5.
8. Janus M, Zeraatkar D, Duku E, et al. Validation of the Early Development Instrument for children with special health needs. *J Paediatr Child Health*. 2019;55(6):659–65. <https://doi.org/10.1111/jpc.14264>
9. Ip P, Li SL, Rao N, et al. Validation study of the Chinese Early Development Instrument (CEDI). *BMC Pediatr*. 2013;13:146. <https://doi.org/10.1186/1471-2431-13-146>
10. D'Angiulli A, Warburton W, Dahinten S, et al. Population-level associations between preschool vulnerability and grade-four basic skills. *PLoS One*. 2009;4(11):e7692. <https://doi.org/10.1371/journal.pone.0007692>
11. Hymel S, Le Mare L, McKee W. The Early Development Instrument (EDI): An examination of convergent and discriminant validity. *Social Indicators Research*. 2011;103(2):267–82.
12. Curtin M, Browne J, Staines A, et al. The Early Development Instrument: an evaluation of its five domains using Rasch analysis. *BMC Pediatr*. 2016;16:10. <https://doi.org/10.1186/s12887-016-0543-8>



13. Brinkman S, Gregory T, Harris J, et al. Associations Between the Early Development Instrument at Age 5, and Reading and Numeracy Skills at Ages 8, 10 and 12: a Prospective Linked Data Study. *Child Indicators Research*. 2013;6(4):695–708. <https://doi.org/10.1007/s12187-013-9189-3>
14. Duku E, Janus M, Brinkman S. Investigation of the Cross-National Equivalence of a Measurement of Early Child Development. *Child Indicators Research*. 2015;8(2):471–89. <https://doi.org/10.1007/s12187-014-9249-3>
15. Guhn M, Gadermann A, Zumbo BD. Does the EDI Measure School Readiness in the Same Way Across Different Groups of Children? *Early Education and Development*. 2007;18(3):453–72. <https://doi.org/10.1080/10409280701610838>
16. Hagquist C, Hellström L. The Psychometric Properties of the Early Development Instrument: A Rasch Analysis Based on Swedish Pilot Data. *Social Indicators Research*. 2014;117(1):301–17. <https://doi.org/10.1007/s11205-013-0344-5>
17. Comaskey B, Roos NP, Brownell M, et al. Maternal depression and anxiety disorders (MDAD) and child development: A Manitoba population-based study. *PLoS One*. 2017;12(5):e0177065. <https://doi.org/10.1371/journal.pone.0177065>
18. Graham MR, Brownell M, Chateau DG, et al. Neurodevelopmental Assessment in Kindergarten in Children Exposed to General Anesthesia before the Age of 4 Years: A Retrospective Matched Cohort Study. *Anesthesiology*. 2016;125(4):667–77. <https://doi.org/10.1097/ALN.0000000000001245>
19. Guhn M, Janus M, Enns J, et al. Examining the social determinants of children's developmental health: protocol for building a pan-Canadian population-based monitoring system for early childhood development. *BMJ Open*. 2016;6(4):e012020. <https://doi.org/10.1136/bmjopen-2016-012020>
20. Janus M, Brinkman S, Duku E, Hertzman C, Santos R, Sayers M, et al. The Early Development Instrument: A population-based measure for communities. A handbook on development, properties, and use. Hamilton, ON: Offord Centre for Child Studies. 2007.
21. Davies S, Janus M, Duku E, et al. Using the Early Development Instrument to examine cognitive and non-cognitive school readiness and elementary student achievement. *Early Childhood Research Quarterly*. 2016;35:63–75. <https://doi.org/10.1016/j.ecresq.2015.10.002>
22. Guhn M, Gadermann AM, Almas A, et al. Associations of teacher-rated social, emotional, and cognitive development in kindergarten to self-reported wellbeing, peer relations, and academic test scores in middle childhood. *Early Childhood Research Quarterly*. 2016;35:76–84. <https://doi.org/10.1016/j.ecresq.2015.12.027>
23. Early Development Instrument Hamilton, Ontario: Offord Centre for Child Studies; 2019 [cited 2020 April 3]. Available from: <https://edi.offordcentre.com/>
24. MOMBABY Database - ICES Intranet Toronto, ON: ICES; 2013 [cited 2020 April 3]. Available from: <https://datadictionary.ices.on.ca/Applications/DataDictionary/Library.aspx?Library=MOMBABY>
25. Chiu M, Lebenbaum M, Lam K, et al. Describing the linkages of the immigration, refugees and citizenship Canada permanent resident data and vital statistics death registry to Ontario's administrative health database. *BMC Med Inform Decis Mak*. 2016;16(1):135. <https://doi.org/10.1186/s12911-016-0375-3>
26. Matheson F, Dunn J, Smith KWL, Moineddin R, Glazier RH. Ontario Marginalization Index user guide. Version 1.0. Centre for Research on Inner City Health, 2012.
27. Drugs and Devices Division, Ontario Public Drug Programs. Understanding Ontario's drug programs: Ministry of Health and Long-term Care; 2020 [cited 2020 April 3]. Available from: <http://www.health.gov.on.ca/en/public/programs/drugs/>
28. Raos R, Janus M. Examining spatial variations in the prevalence of mental health problems among 5-year-old children in Canada. *Social Science & Medicine*. 2011;72(3):383–8. <https://doi.org/10.1016/j.socscimed.2010.09.025>
29. Quick Facts: Ontario Schools, 2014–15: Ontario Ministry of Education; 2016 [cited 2020 April 3]. Available from: [http://www.edu.gov.on.ca/eng/general/elemsec/quickfacts/2014\\_2015.html](http://www.edu.gov.on.ca/eng/general/elemsec/quickfacts/2014_2015.html)
30. Santos R BM, Ekuma O, Mayer T, Soodeen R. . The Early Development Instrument (EDI) in Manitoba: Linking Socioeconomic Adversity and Biological Vulnerability at Birth to Children's Outcomes at Age 5. Winnipeg, MB: Manitoba Centre for Health Policy, 2012.
31. Calam R. Public health implications and risks for children and families resettled after exposure to armed conflict and displacement. *Scand J Public Health*. 2017;45(3):209–11. <https://doi.org/10.1177/1403494816675776>
32. McRae DN, Muhajarine N, Janus M, et al. Immigrant and ethnic neighbourhood concentration and reduced child developmental vulnerability: A Canadian cohort study. *Int J Popul Data Sci*. 2020;5(1):1147. <https://doi.org/10.23889/ijpds.v5i1.1147>
33. Leyton A. A latent class analysis to identify how patterns of Intimate Partner Violence influence Early Childhood Development outcomes in Honduras. *Child Abuse Negl*. 2020;107:104609. <https://doi.org/10.1016/j.chiabu.2020.104609>
34. Jeong J, Adhia A, Bhatia A, et al. Intimate Partner Violence, Maternal and Paternal Parenting, and Early Child Development. *Pediatrics*. 2020;145(6). <https://doi.org/10.1542/peds.2019-2955>



35. Galano MM, Grogan-Kaylor A, Clark HM, et al. Examining the 8-Year Trajectory of Posttraumatic Stress Symptoms in Children Exposed to Intimate Partner Violence. *J Interpers Violence*. 2019;886260519844280. <https://doi.org/10.1177/0886260519844280>
36. The Johns Hopkins ACG System: Johns Hopkins University; [January 15th, 2016]. Available from: <https://acg.jhsph.org/>
37. Austin PC, van Walraven C, Wodchis WP, et al. Using the Johns Hopkins Aggregated Diagnosis Groups (ADGs) to predict mortality in a general adult population cohort in Ontario, Canada. *Med Care*. 2011;49(10):932–9. <https://doi.org/10.1097/MLR.0b013e318215d5e2>
38. Narcotic Monitoring System: Ontario Ministry of Health; 2017 [cited 2020 September 14, 2020]. Available from: [http://www.health.gov.on.ca/en/pro/programs/drugs/ons/monitoring\\_system.aspx](http://www.health.gov.on.ca/en/pro/programs/drugs/ons/monitoring_system.aspx)
39. Gershon AS, Wang C, Guan J, et al. Identifying patients with physician-diagnosed asthma in health administrative databases. *Can Respir J*. 2009;16(6):183–8. <https://doi.org/10.1155/2009/963098>
40. Li B, Quan H, Fong A, et al. Assessing record linkage between health care and Vital Statistics databases using deterministic methods. *BMC Health Serv Res*. 2006;6:48. <https://doi.org/10.1186/1472-6963-6-48>
41. Raffray M, Bayat S, Lassalle M, et al. Linking disease registries and nationwide healthcare administrative databases: the French renal epidemiology and information network (REIN) insight. *BMC Nephrol*. 2020;21(1):25. <https://doi.org/10.1186/s12882-020-1692-4>
42. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885. <https://doi.org/10.1371/journal.pmed.1001885>
43. Adams MM, Wilson HG, Casto DL, et al. Constructing reproductive histories by linking vital records. *Am J Epidemiol*. 1997;145(4):339–48. <https://doi.org/10.1093/oxfordjournals.aje.a009111>
44. Ford JB, Roberts CL, Taylor LK. Characteristics of unmatched maternal and baby records in linked birth records and hospital discharge data. *Paediatr Perinat Epidemiol*. 2006;20(4):329–37. <https://doi.org/10.1111/j.1365-3016.2006.00715.x>
45. Jelleyman T, Spencer N. Residential mobility in childhood and health outcomes: a systematic review. *J Epidemiol Community Health*. 2008;62(7):584–92. <https://doi.org/10.1136/jech.2007.060103>
46. Lawrence E, Root ED, Mollborn S. Residential Mobility in Early Childhood: Household and Neighborhood Characteristics of Movers and Non-Movers. *Demogr Res*. 2015;33:939–50. <https://doi.org/10.4054/demres.2015.33.32>
47. Murphey D, Bandy T, Moore KA. Frequent Residential Mobility and Young Children’s Well-Being. *Child Trends*. 2012; Jan.
48. Wolraich ML, Hagan JF, Jr., Allan C, et al. Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. *Pediatrics*. 2019;144(4). <https://doi.org/10.1542/peds.2019-2528>
49. Kew KM, Carr R, Donovan T, et al. Asthma education for school staff. *Cochrane Database Syst Rev*. 2017;4:CD012255. <https://doi.org/10.1002/14651858.CD012255.pub2>
50. Ouellette-Kuntz HM, Coe H, Lam M, et al. Age at diagnosis of autism spectrum disorders in four regions of Canada. *Can J Public Health*. 2009;100(4):268–73.
51. Chittleborough CR, Searle AK, Smithers LG, et al. How well can poor child development be predicted from early life characteristics?: A whole-of-population data linkage study. *Early Childhood Research Quarterly*. 2016;35:19–30. <https://doi.org/10.1016/j.ecresq.2015.10.006>
52. Neonatal Follow-Up Program Tools and Resources Provincial Council for Maternal and Child Health; 2020 [cited 2020 April 22]. Available from: <https://www.pcmch.on.ca/ontario-neonatal-follow-program/>
53. The Canadian Neonatal Network 2020 [cited 2020 April 22]. Available from: <http://www.canadianneonatalnetwork.org/portal/>
54. Healthy Babies Healthy Children: Ontario Ministry of Children, Community and Social Services; 2020 [cited 2020 April 22]. Available from: <http://www.children.gov.on.ca/htdocs/English/earlychildhood/health/index.aspx>
55. British Columbia’s Early Years Annual Report. British Columbia: Ministry of Children and Family Development, 2015.
56. Zeraatkar D, Duku E, Bennett T, et al. Socioeconomic gradient in the developmental health of Canadian children with disabilities at school entry: a cross-sectional study. *BMJ Open*. 2020;10(4):e032396. <https://doi.org/10.1136/bmjopen-2019-032396>
57. Evans GW, Li D, Whipple SS. Cumulative risk and child development. *Psychol Bull*. 2013;139(6):1342–96. <https://doi.org/10.1037/a0031808>
58. Jutte DP, Brownell M, Roos NP, et al. Rethinking what is important: biologic versus social predictors of childhood health and educational outcomes. *Epidemiology*. 2010;21(3):314–23. <https://doi.org/10.1097/EDE.0b013e3181d61e61>

## Abbreviations

ADHD:	Attention Deficit Hyperactivity Disorder
ASD:	Autism Spectrum Disorder
EDI:	Early Development Instrument
NPV:	Negative predictive value
PPV:	Positive predictive value
RPDB:	Registered Persons Database
SD:	Standardized Differences



## Supplementary Appendix A: Definitions of medical and social risk factors based on health and administrative data

Variable name	Definition	Score
<b>Medical risk factors</b>		
<b>Premature birth</b>	Very preterm: < 28 weeks gestation	2
	Preterm: 28 to ≤ 36 weeks gestation	1
<b>Low birth weight</b>	Very low birth weight: < 1500 g	2
	Low birthweight: 1500 - < 2500 grams	1
<b>ICU at birth</b>	Child spent 3 or more days in a neonatal intensive care nursery during the birth hospitalization.	1
<b>Congenital anomaly</b>	Any one of a list of validated ICD-10-CA diagnosis codes and emergency department or hospitalization discharge record from birth to the child's 4 <sup>th</sup> birthday.	1
<b>Major illness</b>	2 or more major aggregated diagnosis groups (ADG) from birth to the child's 4 <sup>th</sup> birthday. ADG groups considered major are: 3, 9, 11, 12, 13, 18, 25 and 32	1
<b>Hospitalized 6+ ICU stay</b>	Child spent 6 or more days in hospital from birth to the child's 4 <sup>th</sup> birthday.	1
	Child was ever admitted to an intensive care unit from birth discharge date to the child's 4 <sup>th</sup> birthday	1
<b>Social risk factors</b>		
<b>Refugee</b>	Child or mother was an immigrant to Ontario where Immigration category was 'Resettled Refugee'	1
<b>Discharged to social services</b>	Child discharged to social services from birth hospitalization. Information only available for children born in Ontario.	1
<b>Maltreatment</b>	Physician visit, an emergency department visit or hospitalization as a result of intentional maltreatment, adverse circumstances, or violence, from birth to the child's 4 <sup>th</sup> birthday.	1
<b>Low income</b>	Child resided in the lowest neighbourhood quintile based on postal code at the EDI completion date.	1
<b>Income assistance</b>	Child or their mother received prescription drug coverage from the Ontario Drug Benefit Plan at any time between the child's birth date to their 4 <sup>th</sup> birthday.	1
<b>Mother's age at first birth</b>	Mother's age at first delivery < 19 years.	1
<b>Mother's parity status</b>	Mother had 4 or more children as of the child's 4 <sup>th</sup> birthday including the index child.	1
<b>Mother's comorbidity</b>	Mother had 10+ aggregated diagnosis groups (ADG) in the post-natal period (child's birthdate to their 4 <sup>th</sup> birthday) excluding ADG 33 (pregnancy) and ADG 34 (dental).	1
<b>Mother's history of mental illness</b>	Mother had a history of any mental illness (physician visit, ER visit, hospitalization) in the 2-years before the child's birth or from the child's birth to their 4 <sup>th</sup> birthday.	1
<b>Mother's history of assault</b>	Mother had a history of experiencing assault (ER visit or hospitalization) in the 2-years before the child's birth or from the child's birth to their 4 <sup>th</sup> birthday.	1
<b>Other variables used</b>		
<b>Anxiety</b>	A physician office visit, ER visit or hospitalization with a diagnosis of anxiety occurring before or up to two years after the EDI completion date.	NA
<b>Asthma</b>	Any child with one hospitalization with an asthma diagnosis or two physician billing claims with an asthma diagnosis within two years of EDI completion.	NA
<b>Attention deficit hyperactivity disorder</b>	A physician office visit, emergency room visit or hospitalization with a diagnosis of ADHD occurring before or within 2 years of the EDI completion date.	NA
<b>Autism spectrum disorder</b>	A physician office visit, emergency room visit or hospitalization with a diagnosis of ASD occurring before or within 2 years of the EDI completion date.	NA
<b>Cerebral palsy</b>	A physician office visit, emergency room visit, hospitalization with a diagnosis of cerebral palsy occurring before or within 2 years of the EDI completion date.	NA

Supplementary Appendix B: Detailed baseline characteristics of linked children and who completed the EDI and included in study, 2015,  $n = 103,948$

---

**Socio-demographics**


---

Sex	
Female, n (%)	50,308 (48.4)
Male, n (%)	53,640 (51.6)
Age at completion of EDI	
Mean $\pm$ SD	5.68 $\pm$ 0.29
Median (IQR)	6 (5-6)
Place of birth	
Ontario	91,734 (88.2)
Immigrant	3,898 (3.7)
Home Birth	2,854 (2.7)
Inter-provincial migrant	5,462 (5.2)
Rural Residence	
No, n (%)	93,996 (90.4)
Yes, n (%)	9,931 (9.6)
Missing, n (%)	21 (0.0)
Ontario Marginalization Index Material Deprivation Quintile	
Quintile 1 (lowest), n (%)	24,199 (23.3)
Quintile 2, n (%)	22,636 (21.8)
Quintile 3, n (%)	19,314 (18.6)
Quintile 4, n (%)	17,729 (17.1)
Quintile 5, n (%)	19,768 (19.0)
Missing, n (%)	302 (0.3)
Neighbourhood Income Quintile	
Quintile 1 (highest), n (%)	18,885 (18.2)
Quintile 2, n (%)	18,885 (18.2)
Quintile 3, n (%)	21,205 (20.4)
Quintile 4, n (%)	23,015 (22.1)
Quintile 5, n (%)	21,896 (21.1)
Missing, n (%)	62 (0.1)

---

**Medical risk factors**


---

*Prematurity	
Very preterm (< 28 weeks)	218 (0.2)
Preterm (28 to 36 weeks)	6,106 (6.7)
Term (37 weeks +)	85,364 (93.1)
Missing	46 (0.1)
*Birth weight	
Very low birth weight (<1500 g)	606 (0.7)
Low birth weight (<2500 g)	4,035 (4.4)
Normal birth weight ( $\geq$ 2500 g)	87,076 (94.9)
Missing	17 (0.0)
*Neonatal intensive care unit stay	
No, n (%)	86,080 (93.8)
Yes, n (%)	5,654 (6.2)
Missing, (%)	0
Congenital anomalies	
No, n (%)	99,337 (95.6)
Yes, n (%)	2,004 (1.9)
Missing	2,607 (2.5)
ADG minor illness child	
No, n (%)	92,944 (89.4)
Yes, n (%)	8,397 (8.1)
Missing	2,607 (2.5)

---

Continued

## Supplementary Appendix B: Continued

<b>Medical risk factors</b>	
ADG major illness child	
No, n (%)	94,728 (91.1)
Yes, n (%)	6,613 (6.4)
Missing	2,607 (2.5)
Prolonged hospitalization (6+ days)	
No, n (%)	96,861 (93.2)
Yes, n (%)	4,480 (4.3)
Missing	2,607 (2.5)
Any ICU hospitalization	
No, n (%)	99,216 (95.4)
Yes, n (%)	2,125 (2.0)
Missing	2,607 (2.5)
<b>Social risk factors</b>	
*Infant discharged to social services at birth	
No, n (%)	91,656 (99.9)
Yes, n (%)	78 (0.1)
Missing, n (%)	0 (0.0)
*Mother history of pre- or post-natal mental illness	
No, n (%)	47,961 (52.3)
Yes, n (%)	42,251 (46.1)
Missing, n (%)	1,522 (1.7)
*Mother history of assault	
No, n (%)	90,149 (98.2)
Yes, n (%)	63 (0.1)
Missing, n (%)	1,522 (1.7)
*Mother 10+ medical comorbidities	
No, n (%)	53,973 (58.8)
Yes, n (%)	36,239 (39.5)
Missing, n (%)	1,522 (1.7)
Child or mother refugee	
No, n (%)	99,060 (95.3)
Yes, n (%)	3,366 (3.2)
Missing, n (%)	1,522 (1.5)
*Mother's age at first birth < 19 years	
No, n (%)	89,620 (97.6)
Yes, n (%)	4,677 (5.1)
Missing, n (%)	2,114 (2.3)
*Mother with 4+ children	
No, n (%)	83,881 (91.4)
Yes, n (%)	5,731 (6.2)
Missing, n (%)	2,122 (2.3)
History of maltreatment	
No	102,486 (98.6)
Yes	1,462 (1.4)
Missing, n (%)	0 (0.0)
Income Assistance	
No, n (%)	90,732 (87.3)
Yes, n (%)	13,216 (12.7)
Missing, n (%)	0 (0.0)

Continued



## Supplementary Appendix B: Continued

**Health care use**

Annual physician visits	
Mean $\pm$ SD	6.20 $\pm$ 3.74
Median (IQR)	5 (4-8)
Received 18-month enhanced well baby visit	
No, n (%)	54,987 (52.9)
Yes, n (%)	46,354 (44.6)
Missing	2,607 (2.5)

\*Data only available for children born in Ontario as variable based on birth hospitalization record or maternal record linkage with infant birth hospitalization.

Supplementary Appendix C: Number and proportion of children with a linked EDI record with no, at least one, and two or more medical or social risk factors

	Medical risk factors		Social risk factors	
	n	% of children	n	% of children
No risk factors	88,647	85.3	36,488	35.1
At least one risk factor	15,301	14.7	67,460	64.9
Two or more risk factors	9,752	5.3	32,058	34.1

