Heliyon 10 (2024) e31124

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Risk factors for pediatric ischemic stroke and intracranial hemorrhage: A national electronic health record based study

Stuart Fraser [a,c,*,1], Samantha M. Levy [b,1], Amee Moreno [c], Gen Zhu [b], Sean Savitz [c,d], Alicia Zha [e], Hulin Wu [b,c]

[a] Division of Child and Adolescent Neurology, Department of Pediatrics, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA
[b] Department of Biostatistics and Data Science, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA
[c] Institute of Stroke and Cerebrovascular Disease, University of Texas Health Science Center at Houston, Houston, TX, USA
[d] Department of Neurology, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA
[e] Department of Neurology, The Ohio State University College of Medicine, Columbus, OH, USA

A B S T R A C T

Background: Stroke is an important cause of morbidity in pediatrics. Large studies are needed to better understand the epidemiology, pathogenesis and risk factors associated with pediatric stroke. Large administrative datasets can provide information on risk factors in perinatal and childhood stroke at low cost. The aim of this hypothesis-generating study was to use a large administrative dataset to assess for prevalence and odds-ratios of rare exposures associated with pediatric stroke.
Methods: The data for patients aged 0–18 with a diagnosis of either ischemic stroke or intracranial hemorrhage were extracted from the Cerner Health Facts EMR Database from 2000 to 2018. Prevalence of various possible risk factors for pediatric and adult stroke was assessed using ICD 9 and 10 codes. Odds ratios were calculated using a control group of patients without stroke.
Results: 10,688 children were identified with stroke. *6339* (59 %) were ischemic and *4349* (41 %) were hemorrhagic. The most frequently identified risk factors for ischemic stroke across age groups were hypertension (29–44 %), trauma (19–33 %), and malignancy (11–24 %). The most common risk factors seen with hemorrhagic stroke were trauma (32–64 %), malignancy (5–19 %) and arrhythmia (9–12 %). Odds ratios across all age groups for dyslipidemia (17–64), hypertension (20–63), and tobacco exposure (3–59) were high in the ischemic stroke cohort.
Conclusion: This is the largest retrospective study of pediatric stroke of its kind from hospitals across the US in both academic and non-academic clinical settings. Much of our data was consistent with prior studies. ICD codes for tobacco exposure, hyperlipidemia, diabetes, and hypertension all had high odds ratios for stroke in children, which suggest a relationship between these conditions and pediatric stroke. However, ascertainment bias is a major concern with electronic health record-based studies. More focused study is needed into the role of these exposures into the pathogenesis of pediatric stroke.

1. Introduction

Pediatric stroke is a morbid and debilitating disease. It is among the top ten causes of death in children [1,2]. Survivors often have lifelong neurologic deficits including hemiplegia, neurocognitive deficits, and epilepsy [3,4]. The direct medical cost of a single childhood stroke is over $130,000 [5,6]. Indirect cost is higher due to cumulative years of disability and lost income [4].

* Corresponding author. Division of Child and Adolescent Neurology, Department of Pediatrics, McGovern Medical School. University of Texas Health Science Center at Houston, 6410 Fannin Street, Suite 1535, Houston, TX 77030, USA.
E-mail address: Stuart.M.Fraser@uth.tmc.edu (S. Fraser).
[1] Co-first authors, these two authors contributed equally to the manuscript.

https://doi.org/10.1016/j.heliyon.2024.e31124
Received 4 December 2023; Received in revised form 9 May 2024; Accepted 10 May 2024
Available online 11 May 2024

Advancements in understanding pediatric stroke have occurred over the past two decades through the concerted efforts of the International Pediatric Stroke Study and other groups. However, significant knowledge gaps remain concerning pediatric stroke pathogenesis, epidemiology, and outcomes [5,7,8]. Many conditions are frequently associated with pediatric stroke, such as migraine and thrombophilia [9,10]. Large retrospective studies can be used to assess for odds ratios associated with these conditions. Establishing odds ratios of exposures with stroke can be clinically useful and are worth pursuing given our current gaps in knowledge in pediatric stroke.

Large Electronic Health Records (EHR) databases are well suited for studying rare diseases and outcomes as there are a large number of patients available in a single dataset. EHR studies are readily available, have minimal risk of harm to study subjects, and are inexpensive to conduct [11]. We created this hypothesis generating study using a large retrospective EHR database to evaluate exposures associated with pediatric stroke. We had two objectives with this study. The first objective was to qualitatively assess the utility of EHR databases for pediatric stroke research by comparing this EHR data with previous retrospective studies on pediatric stroke. Our second objective was to generate hypotheses in pediatric stroke pathogenesis by assessing for diagnostic associations with a very large cohort of patients. Specifically, we used this large administrative dataset to assess for statistically significant associations with stroke that would be difficult or expensive to find with traditional research methods.

2. Subject and methods

The data used for this project were extracted from the Cerner Health Facts I EMR database (version 2018) in accordance with institutional IRB approval (HSC-MS-18-0124). The Cerner Health Facts® EMR database contains approximately 69 million patients, including 18.7 million patients aged 20 years or younger, across 750 hospitals and clinics nationwide with pediatric patients being seen at 712 of these facilities. This EMR dataset contains data from 2000 to 2018.

2.1. Selection of stroke patients

Patients who had ICD codes for stroke before their 19th birthday were extracted from the Cerner database. Patients with missing age were excluded. Patients who were less than 1 year old but with missing age in days at the time of stroke were also excluded as it is unknown if they would be classified as neonates (<28 days of life) or not. Patients were divided into ischemic stroke and hemorrhagic stroke (defined as nontraumatic intracranial and/or subarachnoid hemorrhage) for all analyses. ICD-9 codes for ischemic stroke were defined as 433.xx, 434.xx, 436.xx, and 437.6x. ICD-10 codes for ischemic stroke were defined as I63.xx, I64.xx, and I676.xx. ICD-9 codes for hemorrhagic stroke were defined as 430.xx, 431.xx, and 432.xx. ICD-10 codes for hemorrhagic stroke were defined as I60.xx, I61.xx, and I62.xx.

2.2. Selection of control subjects

All patients with a visit before 19 years of age or earlier were extracted from the Cerner database. Stroke patients in the Cerner system under 19 years of age were excluded from the control group. Patients were then excluded if they had fewer than 5 diagnoses, including duplicate diagnoses, before 19 years of age. Patients were separated into age groups based on their age at their last recorded encounter.

2.3. Exposure definition

Thirty-three exposures were examined. Exposure definitions are not mutually exclusive. An individual subject may have multiple exposures. Exposures are included if they were diagnosed prior to or concurrently with the first stroke diagnosis. Catheterization, cardiovascular surgery, and neurosurgery were defined based on procedure (CPT codes). The CPT codes used were 93451–93461, 33010–37799, and 61000–62258 for catheterization, cardiovascular surgery, and neurosurgery, respectively. All other exposures were defined by ICD-9 and ICD-10 codes. These codes are specified in Appendix AA.

2.4. Statistical analysis

All analysis was performed using R (R Core Team, 2021). Analyses were stratified by age group and stroke type. Age groups were defined as neonates (up to and less than twenty-eight days of age), twenty-nine days to one year of age, one to four years of age, five to nine years of age, ten to thirteen years of age, and fourteen to eighteen years of age. Age is categorized at the time of first stroke for stroke subjects and time of last recorded diagnosis before the 19th birthday for control subjects. Exposures were summarized by count, proportion, unadjusted odds ratios, and 95 % confidence intervals for unadjusted odds ratios.

Continuous demographic characteristics were summarized using mean, standard deviation, median, and inter-quartile range. Categorical and binary demographic characteristics were summarized using count and proportion. Direct proportions were calculated by dividing the number of subjects in the given age group with the exposure by the total number of subjects in that age group in the sample. Odds ratios were omitted where fewer than 5 subjects in either the stroke or control group had the given exposure. Admission year for control subjects is defined as the year at last encounter before their 19th birthday in the Cerner system. Admission year for pediatric stroke patients is defined as the year at first stroke diagnosis.

Statistical significance is present at a standard 5 % significance level when the 95 % confidence interval does not include zero. An

odds ratio greater than one indicates that the comorbidity might be correlated with an increased risk of stroke. An odds ratio less than one indicates that the comorbidity might be correlated with a decreased risk of stroke.

3. Results

3.1. Study population

We extracted 5,667,386 pediatric patients that met the inclusion criteria. Of this sample, 4349 had hemorrhagic stroke and 6339 had ischemic stroke. 5,656,698 had no diagnosis code for stroke and were included as the control group. Baseline demographic characteristics are presented by type of stroke in Table 1. The highest number of hemorrhagic strokes occur in the 29 days to 1 year old population while the highest number of ischemic strokes occur in the 14- to 18-year-old population. Adjusting for total number of patients in the corresponding age group, the highest incidence of hemorrhagic stroke occurs in the 29 days to 1 year old population with 0.4 % of patients having a hemorrhagic stroke. Similarly, the highest incidence of ischemic stroke occurs in the 29 days to 1 year old population with 0.2 % of patients having an ischemic stroke during this time.

3.2. Comorbidity prevalence

The number and percentage of patients with each comorbidity are summarized by age group for hemorrhagic, ischemic, and non-stroke patients in Table 2, Table 3, and Table 4, respectively. We also present this information graphically in Fig. 1 with a panel for each stroke type (Ischemic stroke, Hemorrhagic stroke, and Control), a color for each age group, and a grouping for each comorbidity. In the hemorrhagic stroke and control group, trauma is the most prevalent comorbidity across all ages except non-stroke neonates. In ischemic stroke patients, hypertension is more prevalent than trauma, while hyperlipidemia, malignancy, and arrhythmia are also highly prevalent across all age groups. For hemorrhagic stroke, zero patients had ICD codes for amphetamine exposure and cocaine use, and these risk factors were removed from the table.

Table 1
EHR database patient demographics (hemorrhagic stroke, ischemic stroke, and control).

		Hemorrhagic Stroke	Ischemic Stroke	Control Group
N		4349	6339	5656698
Age Group	Neonatal, n (%)	274 (6.80)	494 (8.63)	336332 (5.94)
	29 Days–1 year, n (%)	1183 (29.36)	658 (11.49)	312867 (5.52)
	1–4 years, n (%)	830 (20.60)	1278 (22.32)	1,295,770 (22.89)
	5–9 years, n (%)	496 (12.31)	1048 (18.30)	1,199,221 (21.20)
	10–13 years, n (%)	396 (9.83)	774 (13.52)	826038 (14.61)
	14–18 years, n (%)	850 (21.10)	1474 (25.74)	1,686,470 (29.83)
Gender	Male, n (%)	2616 (61.54)	3352 (53.85)	2,948,610 (51.19)
	Female, n (%)	1635 (38.46)	2873 (46.15)	2,811,989 (48.81)
	Missing	*98*	*114*	*74929*
Race	African American, n (%)	840 (21.05)	1166 (20.32)	1,070,534 (20.18)
	Asian, n (%)	84 (2.11)	93 (1.62)	114497 (2.16)
	Hispanic, n (%)	155 (3.88)	167 (2.91)	248893 (4.69)
	White, n (%)	2488 (62.34)	3757 (65.46)	3,291,080 (62.05)
	Other, n (%)	423 (10.60)	556 (9.69)	579147 (10.92)
	Missing	*359*	*600*	*531377*
Admission Year	2000, n (%)	13 (0.30)	37 (0.58)	1462 (0.03)
	2001, n (%)	20 (0.46)	36 (0.57)	6729 (0.12)
	2002, n (%)	29 (0.67)	31 (0.49)	11816 (0.20)
	2003, n (%)	9 (0.21)	20 (0.32)	14658 (0.25)
	2004, n (%)	16 (0.37)	6 (0.09)	13874 (0.24)
	2005, n (%)	21 (0.48)	37 (0.58)	15969 (0.27)
	2006, n (%)	33 (0.76)	70 (1.10)	22969 (0.39)
	2007, n (%)	66 (1.52)	155 (2.45)	37109 (0.64)
	2008, n (%)	108 (2.48)	211 (3.33)	47760 (0.82)
	2009, n (%)	181 (4.16)	343 (5.41)	91343 (1.57)
	2010, n (%)	267 (6.14)	556 (8.77)	161189 (2.76)
	2011, n (%)	350 (8.05)	595 (9.39)	276231 (4.73)
	2012, n (%)	492 (11.31)	689 (10.87)	323670 (5.55)
	2013, n (%)	732 (16.83)	1525 (24.06)	482488 (8.27)
	2014, n (%)	840 (19.31)	1320 (20.82)	624026 (10.69)
	2015, n (%)	754 (17.34)	620 (9.78)	754730 (12.93)
	2016, n (%)	180 (4.14)	50 (0.79)	969420 (16.61)
	2017, n (%)	157 (3.61)	23 (0.36)	1,121,954 (19.23)
	2018, n (%)	81 (1.86)	15 (0.24)	858131 (14.71)

Table 2
Hemorrhagic Stroke Comorbidity Prevalence by Age, n (%).

	Neonatal (N = 274)	29 Days-1 Year (N = 1183)	1–4 Years (N = 830)	5–9 Years (N = 496)	10–13 Years (N = 396)	14–18 Years (N = 850)
Arrhythmia	32 (11.7 %)	111 (9.4 %)	96 (11.6 %)	49 (9.9 %)	40 (10.1 %)	77 (9.1 %)
Cardiac Arrest	3 (1.1 %)	2 (0.2 %)	3 (0.4 %)	2 (0.4 %)	2 (0.5 %)	6 (0.7 %)
Cardiomyopathy	8 (2.9 %)	6 (0.5 %)	7 (0.8 %)	10 (2.0 %)	5 (1.3 %)	11 (1.3 %)
Cerebral Arteritis	0 (0.0 %)	0 (0.0 %)	1 (0.1 %)	0 (0.0 %)	0 (0.0 %)	3 (0.4 %)
Childhood Sepsis	19 (6.9 %)	46 (3.9 %)	34 (4.1 %)	19 (3.8 %)	9 (2.3 %)	21 (2.5 %)
Coagulation Defects	18 (6.6 %)	51 (4.3 %)	34 (4.1 %)	27 (5.4 %)	13 (3.3 %)	38 (4.5 %)
Congenital Heart Disease	34 (12.4 %)	46 (3.9 %)	42 (5.1 %)	15 (3.0 %)	17 (4.3 %)	13 (1.5 %)
Diabetes	2 (0.7 %)	3 (0.3 %)	11 (1.3 %)	11 (2.2 %)	7 (1.8 %)	10 (1.2 %)
Dissection	0 (0.0 %)	0 (0.0 %)	3 (0.4 %)	1 (0.2 %)	2 (0.5 %)	6 (0.7 %)
Heart Failure	4 (1.5 %)	16 (1.4 %)	20 (2.4 %)	11 (2.2 %)	10 (2.5 %)	7 (0.8 %)
HIV/AIDS	0 (0.0 %)	0 (0.0 %)	2 (0.2 %)	1 (0.2 %)	2 (0.5 %)	2 (0.2 %)
Hyperlipidemia	9 (3.3 %)	14 (1.2 %)	30 (3.6 %)	34 (6.9 %)	24 (6.1 %)	28 (3.3 %)
Hypertension	27 (9.9 %)	44 (3.7 %)	75 (9.0 %)	78 (15.7 %)	47 (11.9 %)	90 (10.6 %)
Iron Deficiency Anemia	1 (0.4 %)	16 (1.4 %)	13 (1.6 %)	10 (2.0 %)	5 (1.3 %)	9 (1.1 %)
Malignancy	21 (7.7 %)	62 (5.2 %)	121 (14.6 %)	95 (19.2 %)	50 (12.6 %)	82 (9.7 %)
Meningitis	23 (8.4 %)	27 (2.3 %)	17 (2.1 %)	14 (2.8 %)	8 (2.0 %)	17 (2.0 %)
Migraine	0 (0.0 %)	8 (0.7 %)	13 (1.6 %)	12 (2.4 %)	23 (5.8 %)	44 (5.2 %)
Moyamoya	0 (0.0 %)	1 (0.1 %)	0 (0.0 %)	4 (0.8 %)	5 (1.3 %)	2 (0.2 %)
Neonatal Sepsis	37 (13.5 %)	0 (0.0 %)	2 (0.2 %)	0 (0.0 %)	0 (0.0 %)	0 (0.0 %)
Obesity	6 (2.2 %)	20 (1.7 %)	25 (3.0 %)	25 (5.0 %)	18 (4.6 %)	27 (3.2 %)
PFO	6 (2.2 %)	7 (0.6 %)	10 (1.2 %)	1 (0.2 %)	0 (0.0 %)	1 (0.1 %)
Purpura and Related	21 (7.7 %)	22 (1.9 %)	30 (3.6 %)	26 (5.2 %)	15 (3.8 %)	29 (3.4 %)
Rheumatic	1 (0.4 %)	2 (0.2 %)	11 (1.3 %)	12 (2.4 %)	9 (2.3 %)	14 (1.7 %)
Sickle Cell	1 (0.4 %)	4 (0.3 %)	6 (0.7 %)	5 (1.0 %)	5 (1.3 %)	4 (0.5 %)
Tobacco Exposure	6 (2.2 %)	5 (0.4 %)	23 (2.8 %)	16 (3.2 %)	8 (2.0 %)	57 (6.7 %)
Snoring/Obstructive Sleep Apnea	14 (5.1 %)	20 (1.7 %)	37 (4.5 %)	26 (5.2 %)	9 (2.3 %)	10 (1.2 %)
Trauma	87 (31.8 %)	752 (63.6 %)	463 (55.8 %)	279 (56.3 %)	219 (55.3 %)	501 (58.9 %)
Vascular Malformations	8 (2.8 %)	12 (1.0 %)	14 (1.7 %)	40 (8.1 %)	36 (9.1 %)	63 (7.4 %)
Catheter Angiography	0 (0.0 %)	4 (0.3 %)	3 (0.4 %)	3 (0.6 %)	2 (0.5 %)	2 (0.2 %)
CV Surgery	36 (13.1 %)	91 (7.7 %)	57 (6.9 %)	50 (10.1 %)	26 (6.6 %)	63 (7.4 %)
Neurosurgery	13 (4.7 %)	30 (2.5 %)	22 (2.7 %)	17 (3.4 %)	7 (1.8 %)	23 (2.7 %)

PFO=Patent Foramen Ovale; CV=Cardiovascular.

3.3. Odds ratios

Unadjusted odds ratios with their 95 % confidence intervals are presented in Table 5 and Table 6, respectively. Each column represents an age group, and each row represents a comorbidity. If a value is missing from the table, the value was excluded because too few patients (less than 5) in either the stroke or control group had the comorbidity. Odds ratios and corresponding 95 % confidence intervals are also represented graphically in Fig. 2 for ischemic Stroke and Fig. 3 for hemorrhagic Stroke. In these figures, each panel represents an age group, the point represents the point estimate for the odds ratio, and the line represents the 95 % confidence interval. Odds ratios greater than 140 are displayed as a blue triangle.

Nearly every comorbidity across every age group and stroke type is statistically significant. The odds ratios for dissection and cerebral arteritis are large across all statistically valid age groups for ischemic stroke. The odds ratios for vascular malformations are large across all age groups for hemorrhagic stroke. The odds ratio for snoring and obstructive sleep apnea is particularly large for neonates who had hemorrhagic strokes.

Odds ratios that do not represent statistically significant associations are represented in italics. These include PFO in ischemic stroke patients aged 29 days to 1 year or aged 1–4 years, diabetes and iron deficiency anemia in hemorrhagic stroke patients aged 14–18 years old, and obesity in hemorrhagic stroke patients aged 10–13 years old.

4. Discussion

In this study we analyzed administrative data from hundreds of thousands of pediatric patients in the United States. We found over ten thousand pediatric patients diagnosed with ischemic or hemorrhagic stroke. The results were racially diverse, being 60 % white and 18 % black. Key findings that were consistent with prior validated pediatric stroke research include an increased frequency of stroke in males and that strokes were about 60 % ischemic and 40 % hemorrhagic [12]. We also found a 'U-shaped' incidence of stroke with higher acute ischemic stroke and hemorrhagic stroke incidence in infancy, a nadir in childhood (10–13 years of age), followed by an increase in stroke diagnosis in the teenage years. We found a high prevalence of many well-established comorbidities, including sickle cell, congenital heart disease, and moyamoya in ischemic stroke [13,14]. Nearly all of our assessed risk factors had statistically significant increase in odds ratios compared to controls, likely due to the large number of patients in the database. Ascertainment bias can also affect these finding; many conditions that are highly prevalent in healthy controls, such as PFO, may not be diagnosed in the

Table 3
Ischemic Stroke Comorbidity Prevalence by Age, n (%).

	Neonatal (N = 494)	29 Days-1 Year (N = 658)	1–4 Years (N = 1278)	5–9 Years (N = 1048)	10–13 Years (N = 774)	14–18 Years (N = 1474)
Amphetamine Exposure	0 (0.0 %)	0 (0.0 %)	0 (0.0 %)	1 (0.1 %)	0 (0.0 %)	1 (0.1 %)
Arrhythmia	69 (14.0 %)	119 (18.1 %)	263 (20.6 %)	208 (19.9 %)	145 (18.8 %)	250 (17.0 %)
Cardiac Arrest	1 (0.2 %)	1 (0.2 %)	8 (0.6 %)	3 (0.3 %)	5 (0.7 %)	2 (0.1 %)
Cardiomyopathy	11 (2.2 %)	18 (2.7 %)	45 (3.5 %)	45 (4.3 %)	28 (3.6 %)	41 (2.8 %)
Cerebral Arteritis	1 (0.2 %)	0 (0.0 %)	9 (0.7 %)	3 (0.3 %)	9 (1.2 %)	10 (0.7 %)
Childhood Sepsis	29 (5.9 %)	53 (8.1 %)	57 (4.5 %)	65 (6.2 %)	26 (3.4 %)	64 (4.3 %)
Coagulation Defects	38 (7.7 %)	57 (8.7 %)	101 (7.9 %)	48 (4.6 %)	39 (5.0 %)	88 (6.0 %)
Cocaine Exposure	1 (0.2 %)	1 (0.2 %)	2 (0.2 %)	3 (0.3 %)	0 (0.0 %)	6 (0.4 %)
Congenital Heart Disease	54 (10.9 %)	40 (6.1 %)	71 (5.6 %)	47 (4.5 %)	38 (4.9 %)	42 (2.9 %)
Diabetes	7 (1.4 %)	15 (2.3 %)	75 (5.9 %)	53 (5.1 %)	46 (5.9 %)	45 (3.1 %)
Dissection	1 (0.2 %)	1 (0.2 %)	7 (0.6 %)	11 (1.1 %)	14 (1.8 %)	24 (1.6 %)
Heart Failure	32 (6.5 %)	46 (7.0 %)	127 (9.9 %)	99 (9.5 %)	67 (8.7 %)	84 (5.7 %)
HIV/AIDS	0 (0.0 %)	6 (0.9 %)	20 (1.6 %)	14 (1.3 %)	7 (0.9 %)	12 (0.8 %)
Hyperlipidemia	107 (21.7 %)	132 (20.1 %)	408 (31.9 %)	328 (31.3 %)	224 (28.9 %)	388 (26.3 %)
Hypertension	144 (29.2 %)	192 (29.2 %)	565 (44.2 %)	460 (43.9 %)	294 (38.0 %)	556 (37.7 %)
Iron Deficiency Anemia	12 (2.4 %)	21 (3.2 %)	73 (5.7 %)	45 (4.3 %)	28 (3.6 %)	45 (3.1 %)
Malignancy	54 (10.9 %)	97 (14.7 %)	290 (22.7 %)	254 (24.2 %)	181 (23.4 %)	275 (18.7 %)
Meningitis	20 (4.1 %)	42 (6.4 %)	24 (1.9 %)	24 (2.3 %)	15 (1.9 %)	25 (1.7 %)
Migraine	10 (2.0 %)	20 (3.0 %)	75 (5.9 %)	73 (7.0 %)	69 (8.9 %)	132 (9.0 %)
Moyamoya	0 (0.0 %)	2 (0.3 %)	29 (2.3 %)	59 (5.6 %)	31 (4.0 %)	47 (3.2 %)
Neonatal Sepsis	55 (11.1 %)	0 (0.0 %)	0 (0.0 %)	0 (0.0 %)	1 (0.1 %)	1 (0.1 %)
Obesity	30 (6.1 %)	39 (5.9 %)	125 (9.8 %)	112 (10.7 %)	94 (12.1 %)	136 (9.2 %)
PFO	13 (2.6 %)	9 (1.4 %)	9 (0.7 %)	9 (0.9 %)	8 (1.0 %)	4 (0.3 %)
Purpura and Related	20 (4.1 %)	50 (7.6 %)	76 (6.0 %)	53 (5.1 %)	35 (4.5 %)	68 (4.6 %)
Rheumatic	7 (1.4 %)	20 (3.0 %)	72 (5.6 %)	44 (4.2 %)	39 (5.0 %)	67 (4.6 %)
Sickle Cell	0 (0.0 %)	6 (0.9 %)	37 (2.9 %)	60 (5.7 %)	45 (5.8 %)	83 (5.6 %)
Tobacco Exposure	31 (6.3 %)	37 (5.6 %)	164 (12.8 %)	116 (11.1 %)	76 (9.8 %)	157 (10.7 %)
Snoring/Obstructive Sleep Apnea	27 (5.5 %)	37 (5.6 %)	112 (8.8 %)	86 (8.2 %)	60 (7.8 %)	74 (5.0 %)
Trauma	96 (19.4 %)	199 (30.2 %)	424 (33.2 %)	288 (27.5 %)	245 (31.7 %)	314 (21.3 %)
Vascular Malformations	4 (0.8 %)	11 (1.7 %)	10 (0.8 %)	19 (1.8 %)	8 (1.0 %)	28 (1.9 %)
Catheter Angiography	2 (0.4 %)	1 (0.2 %)	2 (0.2 %)	2 (0.2 %)	3 (0.4 %)	7 (0.5 %)
CV Surgery	45 (9.1 %)	48 (7.3 %)	69 (3.4 %)	87 (8.3 %)	55 (7.1 %)	130 (8.8 %)
Neurosurgery	17 (3.4 %)	19 (2.9 %)	19 (1.5 %)	26 (2.5 %)	21 (2.7 %)	36 (2.4 %)

PFO=Patent Foramen Ovale; CV=Cardiovascular.

medical chart. This could lead to false underestimations of certain exposures in the control group.

4.1. Pediatric ischemic stroke and the odds of risk factors

In our data, we found a relatively high prevalence of many established risk factors for pediatric stroke, such as congenital heart disease (~3–11 % across age groups), malignancy (~11–24 %), coagulation defects (~5–9%), and trauma (~20–33 %). Prior estimates of the prevalence of these risk factors in pediatric ischemic stroke vary widely based on study design and location [5]. Our estimates of prevalence were similar to some prior studies for congenital heart disease [15], malignancy [16], and coagulation defects (including thrombophilia) [17]. Diagnosis of trauma related ICD codes was also common in this dataset, however it had a much weaker association with ischemic stroke when being assessed with odds ratios as compared to most of the other assessed risk factors. We believe this was due to the high prevalence of trauma in the control population. Interestingly, odds ratios for PFO were relatively low (under 10 across all age groups) suggesting that the odds of stroke conferred from the presence of a PFO in childhood is lower than many of the other assessed risk factors. Neonatal stroke also had relatively high odds ratios for snoring/obstructive sleep apnea. We suspect this association is associated more with neonatal encephalopathy than with obstructive sleep apnea as a cause of stroke in this age.

Interestingly, hypertension, diabetes, and hyperlipidemia ICD codes were also frequently present in the ischemic stroke cohort. Trauma is a known risk factor which can cause stroke through direct arterial injury. The role of hypertension and hyperlipidemia is less clear. Prior studies have suggested a possible association between hypertension and childhood ischemic stroke [18]. When we assessed odds ratios for hypertension, hyperlipidemia, and diabetes, the importance of those risk factors was unclear. Diabetes has an odds ratio of 31.1 for childhood stroke in the 1–4 age range group, but lowers to just 3.3 in teenagers. Hyperlipidemia and hypertension have high odds ratios in the neonatal group (62 and 61, respectively), but these odds ratios lower to 17 and 20 in the teenage group. This is contrary to what we would expect based on the cumulative effect of hypertension and hyperlipidemia over a life span [19]. The decrease in magnitude of odds ratio could possibly be attributed to a larger control group in the older pediatric population.

Though the odds ratios are statistically significant, the relatively high prevalence of these traditional 'adult' stroke risk factors such as hypertension, hyperlipidemia, and diabetes should be interpreted cautiously. Ascertainment bias can play a large role in electronic health record studies [20], and it is possible many of the control patients were never tested for hyperlipidemia, hypertension, or diabetes. This could falsely elevate the odds ratio of hypertension and hyperlipidemia in stroke patients. However, the continued

Table 4
Control (Non-Stroke) Comorbidity Prevalence by Age, n (%).

	Neonatal (N = 336332)	29 Days-1 Year (N = 312867)	1–4 Years (N = 1,295,770)	5–9 Years (N = 1,199,221)	10–13 Years (N = 826036)	14–18 Years (N = 1,686,470)
Amphetamine Exposure	4 (0.0 %)	0 (0.0 %)	11 (0.0 %)	12 (0.0 %)	18 (0.0 %)	267 (0.0 %)
Arrhythmia	2049 (0.6 %)	3602 (1.2 %)	12228 (0.9 %)	13639 (1.1 %)	10838 (1.3 %)	36401 (2.2 %)
Cardiac Arrest	68 (0.0 %)	392 (0.1 %)	436 (0.0 %)	181 (0.0 %)	155 (0.0 %)	357 (0.0 %)
Cardiomyopathy	261 (0.1 %)	396 (0.1 %)	1577 (0.1 %)	1657 (0.1 %)	1298 (0.2 %)	3605 (0.2 %)
Cerebral Arteritis	1 (0.0 %)	0 (0.0 %)	5 (0.0 %)	14 (0.0 %)	22 (0.0 %)	79 (0.0 %)
Childhood Sepsis	2093 (0.6 %)	2757 (0.9 %)	6132 (0.5 %)	3826 (0.3 %)	1986 (0.2 %)	5582 (0.3 %)
Coagulation Defects	338 (0.1 %)	563 (0.2 %)	2496 (0.2 %)	3021 (0.3 %)	2273 (0.3 %)	6525 (0.4 %)
Cocaine Exposure	11 (0.0 %)	8 (0.0 %)	51 (0.0 %)	51 (0.0 %)	27 (0.0 %)	888 (0.1 %)
Congenital Heart Disease	7799 (2.3 %)	11448 (3.7 %)	21871 (1.7 %)	12121 (1.0 %)	7397 (0.9 %)	12750 (0.8 %)
Diabetes	219 (0.1 %)	326 (0.1 %)	2590 (0.2 %)	4304 (0.4 %)	6072 (0.7 %)	15734 (0.9 %)
Dissection	0 (0.0 %)	1 (0.0 %)	13 (0.0 %)	12 (0.0 %)	10 (0.0 %)	75 (0.0 %)
Heart Failure	395 (0.1 %)	901 (0.3 %)	3430 (0.3 %)	3095 (0.3 %)	1719 (0.2 %)	3744 (0.2 %)
HIV/AIDS	62 (0.0 %)	113 (0.0 %)	607 (0.1 %)	578 (0.1 %)	367 (0.0 %)	1364 (0.1 %)
Hyperlipidemia	1491 (0.4 %)	1543 (0.5 %)	9473 (0.7 %)	12979 (1.1 %)	13589 (1.7 %)	34317 (2.0 %)
Hypertension	2262 (0.7 %)	2862 (0.9 %)	15988 (1.2 %)	18790 (1.6 %)	14801 (1.78 %)	49532 (2.9 %)
Iron Deficiency Anemia	135 (0.0 %)	777 (0.3 %)	11340 (0.9 %)	8422 (0.7 %)	4061 (0.5 %)	15843 (0.9 %)
Malignancy	3726 (1.1 %)	6542 (2.1 %)	29393 (2.3 %)	30365 (2.5 %)	24624 (3.0 %)	58959 (3.5 %)
Meningitis	458 (0.1 %)	1137 (0.4 %)	3259 (0.3 %)	2429 (0.2 %)	1362 (0.2 %)	2497 (0.2 %)
Migraine	89 (0.0 %)	132 (0.0 %)	1300 (0.10 %)	7025 (0.6 %)	14261 (1.7 %)	51642 (3.1 %)
Moyamoya	0 (0.0 %)	0 (0.0 %)	16 (0.00 %)	51 (0.0 %)	62 (0.0 %)	146 (0.0 %)
Neonatal Sepsis	12288 (3.7 %)	5389 (1.7 %)	11032 (0.85 %)	3717 (0.3 %)	412 (0.1 %)	26 (0.0 %)
Obesity	368 (0.1 %)	531 (0.2 %)	9385 (0.7 %)	29108 (2.4 %)	41738 (5.1 %)	88081 (5.2 %)
PFO	3179 (1.0 %)	5609 (1.8 %)	7037 (0.5 %)	1655 (0.1 %)	876 (0.1 %)	1056 (0.1 %)
Purpura and Related	1087 (0.3 %)	1335 (0.4 %)	6210 (0.5 %)	6742 (0.6 %)	4178 (0.5 %)	9720 (0.6 %)
Rheumatic	156 (0.1 %)	289 (0.1 %)	3153 (0.2 %)	5147 (0.4 %)	4514 (0.6 %)	12865 (0.8 %)
Sickle Cell	169 (0.1 %)	844 (0.3 %)	3476 (0.3 %)	3214 (0.3 %)	2086 (0.3 %)	4455 (0.3 %)
Tobacco Exposure	722 (0.2 %)	473 (0.2 %)	3224 (0.3 %)	3987 (0.3 %)	2869 (0.4 %)	61115 (3.6 %)
Snoring/Obstructive Sleep Apnea	122 (0.0 %)	573 (0.2 %)	15063 (1.2 %)	26354 (2.2 %)	13951 (1.7 %)	16227 (1.0 %)
Trauma	3459 (1.0 %)	28249 (9.0 %)	387653 (29.9 %)	491676 (41.0 %)	391484 (47.4 %)	774394 (45.9 %)
Vascular Malformations	29 (0.0 %)	73 (0.0 %)	402 (0.0 %)	436 (0.0 %)	381 (0.0 %)	778 (0.0 %)
Catheter Angiography	4 (0.0 %)	12 (0.0 %)	128 (0.0 %)	113 (0.0 %)	63 (0.0 %)	218 (0.01 %)
CV Surgery	3938 (1.2 %)	5350 (1.7 %)	28776 (2.2 %)	25410 (2.1 %)	17382 (2.1 %)	53877 (3.2 %)
Neurosurgery	15 (0.0 %)	294 (0.1 %)	1771 (0.1 %)	1538 (0.1 %)	1035 (0.1 %)	1980 (0.1 %)

PFO=Patent Foramen Ovale; CV=Cardiovascular.

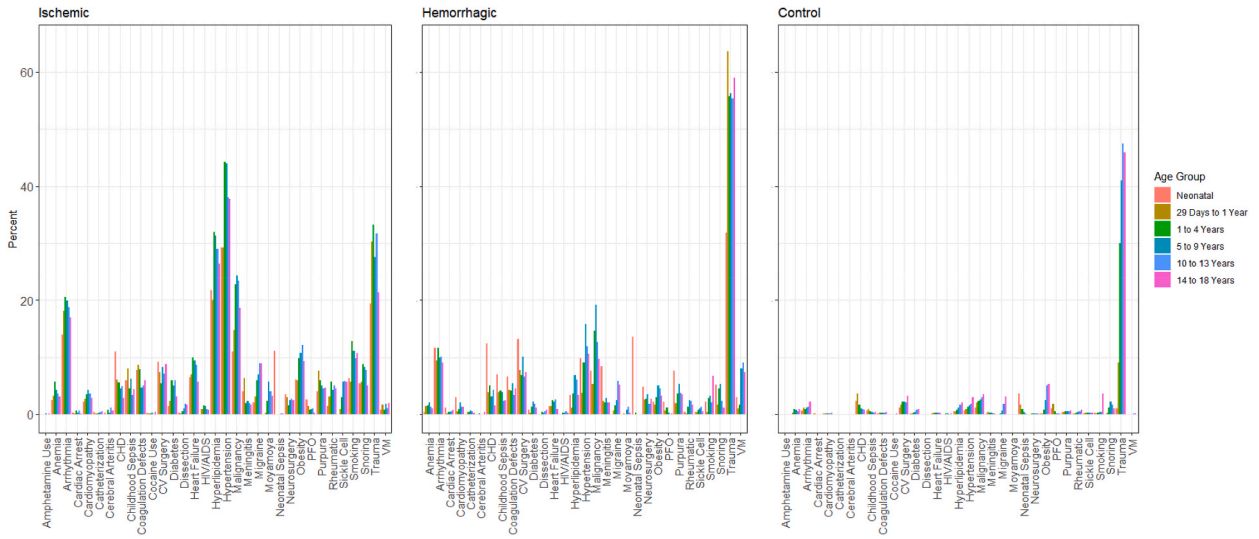

Fig. 1. Percent Prevalence of Different Comorbidities by Stroke Type and Age. The panel on the far left demonstrates prevalence percentage by ischemic stroke for different age groups. The middle panel demonstrates prevalence percentage for hemorrhagic stroke for different age groups, and the control group's comorbidity prevalence by age group is shown on the right.
(Abbreviations: CHD=Congestive Heart Failure, CV=Cardiovascular, PFO=Patent Foramen Ovale; VM= Vascular).

Table 5
Ischemic stroke odds ratios and confidence intervals by age.

	Neonatal	29 Days-1 Year	1–4 Years	5–9 Years	10–13 Years	14–18 Years
Amphetamine Use						
Arrhythmia	26.5 (20.5,34.3)	19.0 (15.5,23.2)	27.2 (23.7,31.2)	21.5 (18.5,25.1)	17.3 (14.5,20.8)	9.3 (8.1,10.6)
Cardiac Arrest			18.7 (9.3,37.7)		34.6 (14.2,84.7)	
Cardiomyopathy	29.3 (15.9,54.0)	22.2 (13.8,35.8)	30.0 (22.2,40.5)	32.4 (24.0,43.9)	23.9 (16.3,34.9)	13.4 (9.8,18.3)
Cerebral Arteritis			1838.0 (615.1,5491.8)		441.7 (202.7962.4)	145.8 (75.4282.1)
Childhood Sepsis	10.0 (6.8,14.5)	9.9 (7.4,13.1)	9.8 (7.5,12.8)	20.7 (16.0,26.6)	14.4 (9.7,21.4)	13.7 (10.6,17.6)
Coagulation Defects	82.8 (58.5117.3)	52.6 (39.6,69.9)	44.5 (36.2,54.7)	19.0 (14.2,25.5)	19.2 (13.9,26.6)	16.4 (13.2,20.3)
Cocaine Use						7.8 (3.5,17.3)
Congenital Heart Disease	5.2 (3.9,6.9)	1.7 (1.2,2.4)	3.4 (2.7,4.4)	4.6 (3.4,6.2)	5.7 (4.1,7.9)	3.9 (2.8,5.2)
Diabetes	22.1 (10.3,47.1)	22.4 (13.3,37.7)	31.1 (24.6,39.4)	14.8 (11.2,19.5)	8.5 (6.3,11.5)	3.3 (2.5,4.5)
Dissection			549.0 (218.7,1378.2)	1060.1 (466.7,2407.8)	1521.6 (673.8,3436.3)	372.2 (234.4591.0)
Heart Failure	58.9 (40.6,85.4)	26.0 (19.2,35.4)	41.6 (34.5,50.1)	40.3 (32.7,49.7)	45.4 (35.2,58.6)	27.2 (21.7,33.9)
HIV/AIDS		25.5 (11.2,58.1)	33.9 (21.7,53.1)	28.1 (16.5,47.9)	20.5 (9.7,43.5)	10.1 (5.7,17.9)
Hyperlipidemia	62.1 (49.8,77.4)	50.6 (41.6,61.7)	63.7 (56.5,71.8)	41.6 (36.5,47.5)	24.4 (20.8,28.5)	17.2 (15.3,19.3)
Hypertension	60.8 (49.8,74.1)	44.6 (37.6,53.0)	63.4 (56.7,70.9)	49.2 (43.5,55.6)	33.6 (29.0,38.9)	20.0 (18.0,22.3)
Iron Deficiency Anemia	62.0 (34.1112.7)	13.2 (8.5,20.6)	6.9 (5.4,8.7)	6.34 (4.7,8.6)	7.6 (5.2,11.1)	3.3 (2.5,4.5)
Malignancy	11.0 (8.2,14.6)	8.1 (6.5,10.1)	12.7 (11.1,14.4)	12.3 (10.7,14.2)	9.9 (8.4,11.7)	6.3 (5.6,7.2)
Meningitis	30.9 (19.6,48.9)	18.7 (13.6,25.7)	7.6 (5.1,11.4)	11.6 (7.7,17.3)	12.0 (7.2,20.0)	11.6 (7.8,17.3)
Migraine	78.1 (40.4151.0)	74.3 (46.1119.6)	62.1 (48.9,78.9)	12.7 (10.0,16.1)	5.6 (4.4,7.1)	3.1 (2.6,3.7)
Moyamoya			1880.4 (1018.7,3470.7)	1402.7 (959.4,2050.9)	555.8 (359.0,860.6)	380.4 (272.7530.6)
Neonatal Sepsis	3.3 (2.5,4.4)					
Obesity	59.0 (40.2,86.6)	37.1 (26.5,51.8)	14.9 (12.3,17.9)	4.8 (4.0,5.9)	2.6 (2.1,3.2)	1.8 (1.6,2.2)
PFO	2.8 (1.6,4.9)	*0.8 (0.4,1.5)*	*1.3 (0.7,2.5)*	6.3 (3.3,12.1)	9.8 (4.9,19.8)	
Purpura and Related	13.0 (8.3,20.4)	19.2 (14.3,25.7)	13.1 (10.4,16.6)	9.4 (7.1,12.4)	9.3 (6.6,13.1)	8.3 (6.5,10.7)
Rheumatic	31.0 (14.5,66.4)	33.9 (21.4,53.7)	24.5 (19.3,31.1)	10.2 (7.5,13.8)	9.7 (7.0,13.3)	6.12 (4.9,7.9)
Sickle Cell		3.4 (1.5,7.6)	11.1 (8.0,15.4)	22.6 (17.4,29.4)	24.4 (18.0,33.1)	22.5 (18.0,28.2)
Tobacco Exposure	31.1 (21.5,45.1)	39.4 (27.9,55.5)	59.0 (49.9,69.8)	37.3 (30.7,45.4)	31.2 (24.6,39.7)	3.2 (2.7,3.7)
Snoring/Obstructive Sleep Apnea	159.3 (104.0,244.1)	32.5 (23.1,45.7)	8.2 (6.7,9.9)	4.0 (3.2,5.0)	4.9 (3.8,6.4)	5.4 (4.3,6.9)
Trauma	23.21 (18.5,29.1)	4.4 (3.7,5.2)	1.2 (1.0,1.3)	0.6 (0.5,0.6)	0.5 (0.4,0.6)	0.3 (0.3,0.4)
Vascular Malformations		72.7 (38.4137.8)	25.4 (13.5,47.7)	50.8 (32.0,80.8)	22.7 (11.2,45.8)	42.0 (28.7,61.5)
Catheter Angiography						36.9 (17.4,78.5)
CV Surgery	8.5 (6.2,11.5)	4.5 (3.4,6.1)	2.5 (2.0,3.2)	4.2 (3.4,5.2)	3.6 (2.7,4.7)	2.9 (2.5,3.5)
Neurosurgery	799.1 (396.8,1609.3)	31.6 (19.8,50.6)	11.0 (7.0,17.4)	19.8 (13.4,29.3)	22.2 (14.4,34.4)	21.3 (15.3,29.7)

PFO=Patent Foramen Ovale; CV=Cardiovascular.

implication of hypertension and dyslipidemia in pediatric arterial ischemic stroke remains an important topic of investigation. This is the second time this association has been found in a large electronic health record study. It was reported in an ICD based search of a national EHR database by Lo and colleagues in 2009 [21]. Additionally, the increased prevalence of dyslipidemia and hypertension in pediatric stroke has been recognized in prior retrospective studies [22,23].

Tobacco exposure ICD codes were also associated with an increased odds of ischemic stroke across all age groups. This is a novel finding to our knowledge which has not been previously reported. The pathophysiologic implications of this are unclear. The known relative risk of stroke is greatly increased in smoking adults when compared to nonsmoking adults [24], and childhood second hand smoke exposure has been linked with adult stroke in the past [25]. However, just as with lipid screening and diabetes, ascertainment bias complicates the interpretation of these findings. Many children with exposure to tobacco in the control group may not have had this coded for in their medical chart. Still, these findings generate an interesting question about the role of environmental exposure in pediatric stroke and suggest that exposure to tobacco may increase stroke risk in children. However, further research is necessary

Table 6
Hemorrhagic stroke odds ratios and confidence intervals by age.

	Neonatal	29 Days-1 Year	1–4 Years	5–9 Years	10–13 Years	14–18 Years
Arrhythmia	21.6 (14.9,31.2)	8.9 (7.3,10.8)	13.7 (11.1,17.0)	9.5 (7.1,12.8)	8.5 (6.1,11.7)	4.5 (3.6,5.7)
Cardiac Arrest						33.6 (14.9,75.5)
Cardiomyopathy	38.7 (19.0,79.1)	4.0 (1.8,9.0)	7.0 (3.3,14.7)	14.9 (7.9,27.9)	8.1 (3.4,19.7)	6.1 (3.4,11.1)
Childhood Sepsis	11.9 (7.5,19.0)	4.6 (3.4,6.1)	9.0 (6.4,12.7)	12.5 (7.9,19.7)	9.7 (5.0,18.7)	7.6 (4.9,11.8)
Coagulation Defects	69.9 (42.8114.1)	25.0 (18.7,33.5)	22.1 (15.7,31.3)	22.8 (15.4,33.7)	12.3 (7.1,21.4)	12.1 (8.7,16.7)
Congenital Heart Disease	6.0 (4.2,8.6)	*1.1 (0.8,1.4)*	3.1 (2.3,4.2)	3.1 (1.8,5.1)	5.0 (3.1,8.1)	2.0 (1.2,3.5)
Diabetes			6.7 (3.7,12.2)	6.3 (3.5,11.5)	2.4 (1.2,5.1)	*1.3 (0.7,2.4)*
Dissection						159.9 (69.4368.2)
Heart Failure		4.8 (2.9,7.8)	9.3 (6.0,14.5)	8.8 (4.8,16.0)	12.4 (6.6,23.3)	3.7 (1.8,7.9)
HIV/AIDS						
Hyperlipidemia	7.6 (3.9,14.9)	2.4 (1.4,4.1)	5.1 (3.5,7.3)	6.7 (4.8,9.5)	3.9 (2.6,5.8)	1.6 (1.1,2.4)
Hypertension	16.1 (10.8,24.1)	4.2 (3.1,5.7)	8.0 (6.3,10.1)	11.7 (9.2,14.9)	7.4 (5.4,10.0)	3.9 (3.1,4.9)
Iron Deficiency Anemia		5.5 (3.4,9.1)	1.8 (1.0,3.1)	2.9 (1.6,5.4)	2.6 (1.1,6.3)	*1.1 (0.6,2.2)*
Malignancy	7.4 (4.7,11.6)	2.6 (2.0,3.4)	7.4 (6.1,8.9)	9.1 (7.3,11.4)	4.7 (3.5,6.3)	2.95 (2.35,3.7)
Meningitis	67.2 (43.4104.0)	6.4 (4.4,9.4)	8.3 (5.1,13.4)	14.3 (8.4,24.4)	12.5 (6.2,25.2)	13.8 (8.5,22.3)
Migraine		16.1 (7.9,33.0)	15.8 (9.1,27.5)	4.2 (2.4,7.5)	3.5 (2.3,5.4)	1.7 (1.3,2.3)
Neonatal Sepsis	4.1 (2.9,5.8)					
Obesity	20.4 (9.0,46.2)	10.1 (6.5,15.9)	4.3 (2.9,6.3)	2.1 (1.4,3.2)	*0.9 (0.6,1.4)*	0.6 (0.4,0.9)
PFO	2.4 (1.0,5.3)	0.3 (0.2,0.7)	2.2 (1.2,4.2)			
Purpura and Related	25.6 (16.3,40.1)	4.4 (2.9,6.8)	7.8 (5.4,11.2)	9.8 (6.6,14.5)	7.7 (4.6,13.0)	6.1 (4.2,8.8)
Rheumatic			5.5 (3.0,10.0)	5.8 (3.2,10.2)	4.2 (2.2,8.2)	2.2 (1.3,3.7)
Sickle Cell			2.7 (1.2,6.1)	3.8 (1.2,9.2)	5.1 (2.1,12.2)	
Tobacco Exposure	10.4 (4.6,23.5)	2.8 (1.2,6.8)	11.4 (7.5,17.3)	10.0 (6.1,16.5)	5.9 (2.9,11.9)	1.9 (1.5,2.5)
Snoring/Obstructive Sleep Apnea	148.4 (84.2261.4)	9.4 (6.0,14.7)	4.0 (2.9,5.5)	2.5 (1.7,3.7)	1.4 (0.7,2.6)	1.2 (0.7,2.3)
Trauma	44.8 (34.6,57.9)	17.6 (15.6,19.8)	3.0 (2.6,3.4)	1.9 (1.6,2.2)	1.4 (1.1,1.7)	1.7 (1.5,1.9)
Vascular Malformations	348.8 (158.0, 770.0)	43.8 (23.7, 81.0)	55.3 (32.3,94.6)	241.4 (172.4338.1)	217.1 (151.9310.2)	173.7 (133.1226.6)
Catheter Angiography						
CV Surgery	12.8 (9.0,18.2)	4.8 (3.9,5.9)	3.3 (2.5,4.3)	5.2 (3.9,6.9)	3.3 (2.2,4.9)	2.4 (1.9,3.1)
Neurosurgery	1116.8 (526.2,2370.2)	27.7 (18.9,40.5)	19.9 (13.0,30.5)	27.6 (17.0,45.0)	14.3 (6.8,30.4)	23.7 (15.6,35.9)

PFO=Patent Foramen Ovale; CV=Cardiovascular.

before drawing any definitive conclusions regarding hypertension, hyperlipidemia, diabetes, and tobacco exposure in pediatric stroke. Specifically, prospective research with both healthy pediatric controls and hospitalized pediatric stroke patients looking at dyslipidemia, hypertension, and tobacco exposure could provide important information on their role in stroke pathogenesis.

4.2. Intracerebral hemorrhage and the odds of risk factors

Spontaneous intracerebral hemorrhage is most commonly attributable to vascular malformations in children. In this study, vascular malformations using the ICD codes we assessed were present in less than 15 % of patients across all age groups of hemorrhagic stroke patients. Interestingly, this was *also* noted in Lo et al.'s study in 2009 using the KIDS inpatient database [21]. Given the results of previous retrospective studies on pediatric hemorrhagic stroke, we expect that this is due to arteriovenous malformations being under-coded for in EHR databases [12,26]. Regarding the prevalence and odds ratios of other risk factors, comorbid ICD codes for trauma were extremely prevalent in this dataset, ranging from ~30 to 60 % in different age ranges. These prevalence values were very similar to the control group in childhood and teenage patients. Odds ratios for trauma and intracranial hemorrhage were high in younger age groups (OR of 40 for neonates) but decreased to nearly 1 by the teenage years. Similar to results in the ischemic stroke cohort, this could be the result of inaccurate ICD codes or an indication of a true decrease in odds. We cannot discern the difference based on this study design.

In this study, procedures and conditions associated with surgery and critical illness in early life generally had high odds ratios of intracranial hemorrhage in the neonatal period and infancy, including congenital heart disease, neonatal sepsis, and heart disease. Coagulation defects, cardiac disease, and purpura and related diseases continued to have odds ratios above 20 into childhood, suggesting an association of these diseases with an increased risk of hemorrhagic stroke.

Hypertension, hyperlipidemia, and tobacco exposure did have statistically significant odds ratios in association with hemorrhagic stroke. Hypertension odds ratios ranged from 3.9 to 16.1, hyperlipidemia from 1.6 to 7.6, and tobacco exposure odds ratios for ICH ranged from 1.9 to 11.4. These values were all lower than the odds ratios associated with ischemic stroke. The significance of these findings is not clear in this study. In our opinion, it does again raise the interesting question about the pathological link between hypertension, dyslipidemia, tobacco exposure and ischemic stroke in childhood.

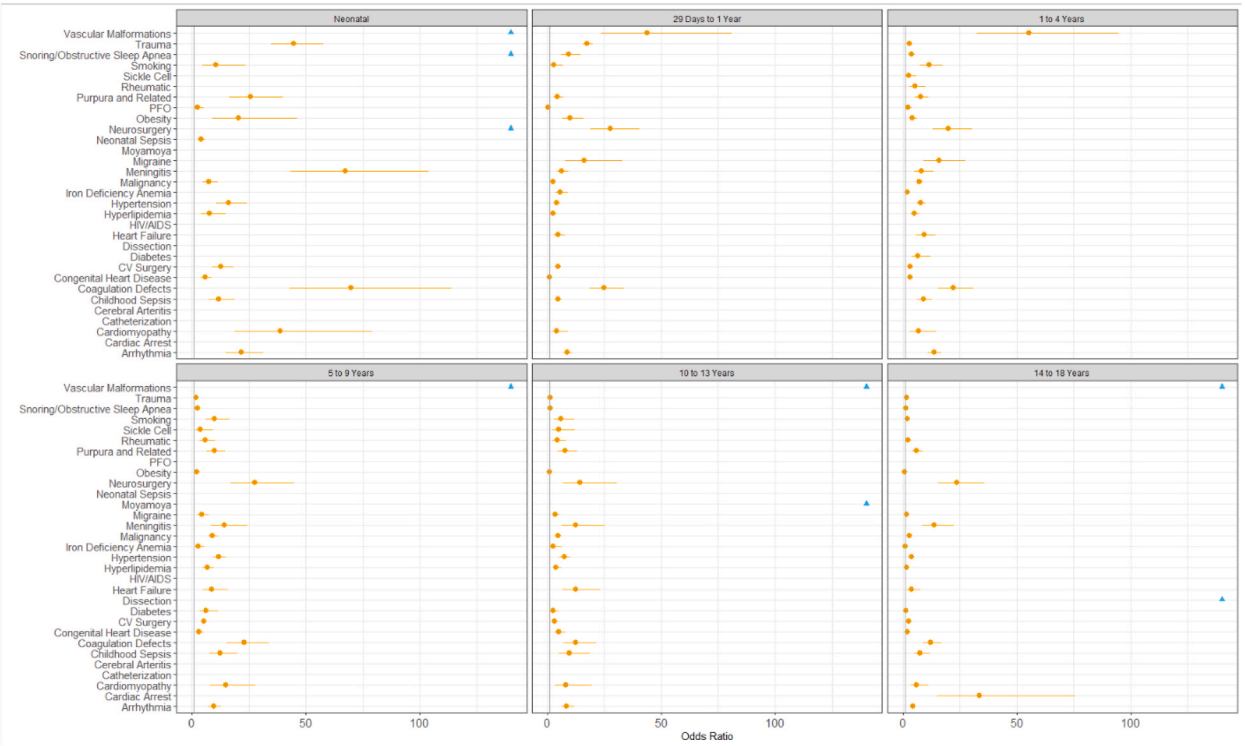

Fig. 2. Ischemic Stroke Odds Ratios of Different Comorbidities Divided by Age. Panels are labeled according to age groups as indicated by the label at the top of each panel. A blue triangle indicates an odds ratio greater than 150 for the listed comorbidity.

Fig. 3. Hemorrhagic Stroke Odds Ratios of Different Comorbidities Divided by Age. Panels are labeled according to age groups as indicated by the label on the top of each panel. A blue triangle indicates an odds ratio greater than 150.

4.3. Limitations and further work

There are important limitations in using EHR data to study pediatric stroke. Type of stroke was determined by the age at which the first stroke diagnosis code appeared. Unfortunately, it is possible that some presumed perinatal stroke patients will receive their first ICD diagnosis of stroke at 6 months of age or older, which will may have led to some presumed perinatal stroke patients being analyzed as childhood-onset stroke in this study. We did not include specific ICD codes for presumed perinatal stroke, however, which we believe will mitigate some of the risk of inclusion of presumed perinatal stroke patients. Additionally, arteriopathy, which is implicated in up to 50 % of pediatric stroke, seems to be under-coded in this EHR data. Arteritis and dissection were present in less than 1 % of pediatric arterial ischemic stroke patients in our study, which is likely far, far below the actual incidence in this cohort.

A recent study from the Swedish national registry for pediatric acute ischemic stroke using ICD-10 codes carried a positive predictive value of 89 % [27]. ICD-9 codes have been found to have positive predictive values for pediatric intracranial hemorrhage of 82 %, but there are no validation studies for ICD-10 codes for pediatric intracranial hemorrhage and subarachnoid hemorrhage to our knowledge [28]. Adult intracranial hemorrhage and subarachnoid hemorrhage is well studied using EHR data in adult literature, with positive predictive values of ICD codes >85 % for nontraumatic intracranial hemorrhage and subarachnoid hemorhage [29]. Given these prior findings, we estimate that 80–90 % of the stroke cases we identified were truly stroke. It is also possible that cases of stroke were missed using this system to identify cases, but we are unable to quantify this number without conducting a full validation study. On a smaller scale, validation studies should be performed to develop better ways to identify and confirm pediatric stroke ICD codes from EHR data in the future.

5. Conclusion

This hypothesis generating EHR study, the largest of its kind, assessed a racially and geographically diverse pediatric patient population for risk factors associated with ischemic and hemorrhagic stroke in both inpatient and outpatient centers. Hospitals in this study were both academic and non-academic, and included outpatient clinics, dedicated children's hospitals and adult hospitals with pediatric wards. Within the limitations of this design, the novel findings of this large retrospective study are intriguing. The role of many 'non-classic' pediatric stroke risk factors, such as hypertension and dyslipidemia raise the question of the role of these conditions in pediatric stroke and suggest the need for further study of these risk factors going forward. The high prevalence of ICD codes related to tobacco exposure in pediatric ischemic stroke, in particular, was a novel finding in this study. An understanding of the relative risk of stroke inferred by these conditions in children will be important for long term secondary prevention strategies in the future.

Data availability

The data used for this project were extracted from the Cerner Health Facts I EMR database (version 2018). Derived data presented in this study are available from Dr. Hulin Wu upon request.

CRediT authorship contribution statement

Stuart Fraser: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization. **Samantha M. Levy:** Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Amee Moreno:** Writing – review & editing, Formal analysis, Conceptualization. **Gen Zhu:** Writing – review & editing, Formal analysis, Data curation. **Sean Savitz:** Writing – review & editing, Supervision, Conceptualization. **Alicia Zha:** Supervision, Conceptualization. **Hulin Wu:** Writing – review & editing, Writing – original draft, Supervision, Software, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This project is partially supported by the Center for Big Data in Health Sciences (CBD-HS) at the School of Public Health, University of Texas Health Science Center at Houston. This work was also possible due to funding from the National Institute of Health Training Grant 5T32NS007412-23. The authors appreciate the discussions, suggestions, and inputs from Chenguang Zhang, Hongjian Zhu, and other Stroke EHR Working Group members.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e31124.

References

[1] R.v. Krishnamurthi, G. deVeber, V.L. Feigin, et al., Stroke prevalence, Mortality and disability-adjusted life years in children and youth aged 0-19 Years: data from the global and regional burden of stroke 2013, Neuroepidemiology 45 (3) (2015) 177–189, https://doi.org/10.1159/000441087.

[2] 10 leading causes of death by age group, National Center for Injury Prevention and Control, CDC, 2018, https://www.cdc.gov/injury/wisqars/pdf/leading_causes_of_death_by_age_group_2015-a. (Accessed 10 May 2022).

[3] A. Fuentes, A. Deotto, M. Desrocher, G. DeVeber, R. Westmacott, Determinants of cognitive outcomes of perinatal and childhood stroke: a review, Child Neuropsychol. 22 (1) (2016) 1–38, https://doi.org/10.1080/09297049.2014.969694.

[4] C. Ellis, K. McGrattan, P. Mauldin, B. Ovbiagele, Costs of pediatric stroke care in the United States: a systematic and contemporary review, Expert Rev. Pharmacoecon. Outcomes Res. 14 (5) (2014) 643–650, https://doi.org/10.1586/14737167.2014.933672.

[5] D.M. Ferriero, H.J. Fullerton, T.J. Bernard, et al., Management of stroke in neonates and children: a scientific statement from the American heart association/American stroke association, Stroke 50 (3) (2019) e51–e96, https://doi.org/10.1161/STR.0000000000000183.

[6] M.A. Gardner, N.K. Hills, S. Sidney, S.C. Johnston, H.J. Fullerton, The 5-year direct medical cost of neonatal and childhood stroke in a population-based cohort, Neurology 74 (5) (2010) 372–378, https://doi.org/10.1212/WNL.0b013e3181cbcd48.

[7] R.J. Felling, M.F. Rafay, T.J. Bernard, et al., Predicting recovery and outcome after pediatric stroke: results from the international pediatric stroke study, Ann. Neurol. 87 (6) (2020) 840–852, https://doi.org/10.1002/ana.25718.

[8] M. Greenham, A. Gordon, V. Anderson, M.T. Mackay, Outcome in childhood stroke, Stroke 47 (4) (2016) 1159–1164, https://doi.org/10.1161/STROKEAHA.115.011622.

[9] C.C. Trenor, A.D. Michelson, Thrombophilia and pediatric stroke, Circulation 121 (16) (2010) 1795–1797, https://doi.org/10.1161/CIR.0b013e3181df6410.

[10] A.A. Gelfand, H.J. Fullerton, A. Jacobson, et al., Is migraine a risk factor for pediatric stroke? Cephalalgia 35 (14) (2015) 1252–1260, https://doi.org/10.1177/0333102415576222.

[11] J.A. Casey, B.S. Schwartz, W.F. Stewart, N.E. Adler, Using electronic health records for population health research: a review of methods and applications, Annu Rev Public Health 37 (1) (2016) 61–81, https://doi.org/10.1146/annurev-publhealth-032315-021353.

[12] G. Boulouis, T. Blauwblomme, J.F. Hak, et al., Nontraumatic pediatric intracerebral hemorrhage, Stroke 50 (12) (2019) 3654–3661, https://doi.org/10.1161/STROKEAHA.119.025783.

[13] H.J. Fullerton, Y.W. Wu, S. Zhao, S.C. Johnston, Risk of stroke in children: ethnic and gender disparities, Neurology 61 (2) (2003) 189–194, https://doi.org/10.1212/01.WNL.0000078894.79866.95.

[14] J. Pappachan, F.J. Kirkham, Cerebrovascular disease and stroke, Arch. Dis. Child. 93 (10) (2008) 890–898, https://doi.org/10.1136/adc.2008.142836.

[15] C.K. Fox, S. Sidney, H.J. Fullerton, Community-based case–control study of childhood stroke risk associated with congenital heart disease, Stroke 46 (2) (2015) 336–340, https://doi.org/10.1161/STROKEAHA.114.007218.

[16] L.R. Sun, A. Linds, M. Sharma, et al., Cancer and tumor-associated childhood stroke: results from the international pediatric stroke study, Pediatr. Neurol. 111 (2020) 59–65, https://doi.org/10.1016/j.pediatrneurol.2020.06.002.

[17] R. Zadro, D. Coen Herak, Inherited prothrombotic risk factors in children with first ischemic stroke, Biochem. Med. (2012) 298–310, https://doi.org/10.11613/BM.2012.033. Published online.

[18] J.C. Kupferman, D.I. Zafeiriou, M.B. Lande, F.J. Kirkham, S.G. Pavlakis, Stroke and hypertension in children and adolescents, J. Child Neurol. 32 (4) (2017) 408–417, https://doi.org/10.1177/0883073816685240.

[19] A. Daugherty, A.R. Tall, Daemen Mjap, et al., Recommendation on design, execution, and reporting of animal atherosclerosis studies: a scientific statement from the American heart association, Arterioscler. Thromb. Vasc. Biol. 37 (9) (2017), https://doi.org/10.1161/ATV.0000000000000062.

[20] S. Khurshid, C. Reeder, L.X. Harrington, et al., Cohort design and natural language processing to reduce bias in electronic health records research, NPJ Digit Med 5 (1) (2022) 47, https://doi.org/10.1038/s41746-022-00590-0.

[21] W. Lo, J. Stephens, S. Fernandez, Pediatric stroke in the United States and the impact of risk factors, J. Child Neurol. 24 (2) (2009) 194–203, https://doi.org/10.1177/0883073808322665.

[22] S. Sultan, M. Dowling, A. Kirton, et al., Dyslipidemia in children with arterial ischemic stroke: prevalence and risk factors, Pediatr. Neurol. 78 (2018) 46–54, https://doi.org/10.1016/j.pediatrneurol.2017.09.019.

[23] M.M. Adil, L.A. Beslow, A.I. Qureshi, A.A. Malik, L.C. Jordan, Hypertension is associated with increased mortality in children hospitalized with arterial ischemic stroke, Pediatr. Neurol. 56 (2016) 25–29, https://doi.org/10.1016/j.pediatrneurol.2015.11.002.

[24] B. Pan, X. Jin, L. Jun, S. Qiu, Q. Zheng, M. Pan, The relationship between smoking and stroke, Medicine 98 (12) (2019) e14872, https://doi.org/10.1097/MD.0000000000014872.

[25] M. Pistilli, V.J. Howard, M.M. Safford, et al., Association of secondhand tobacco smoke exposure during childhood on adult cardiovascular disease risk among never-smokers, Ann. Epidemiol. 32 (2019) 28–34.e1, https://doi.org/10.1016/j.annepidem.2019.01.012.

[26] G. Boulouis, S. Stricker, S. Benichi, et al., Etiology of intracerebral hemorrhage in children: cohort study, systematic review, and meta-analysis, J. Neurosurg. Pediatr. 27 (3) (2021) 357–363, https://doi.org/10.3171/2020.7.PEDS20447.

[27] A. Walås, K. Svensson, M. Gyris, P. Bang, H.E.K. Sundelin, Paediatric ischaemic stroke is a valid diagnosis in the Swedish National Patient Register, Acta Paediatr. 110 (7) (2021) 2179–2186, https://doi.org/10.1111/apa.15791.

[28] N. Agrawal, S.C. Johnston, Y.W. Wu, S. Sidney, H.J. Fullerton, Imaging data reveal a higher pediatric stroke incidence than prior US estimates, Stroke 40 (11) (2009) 3415–3421, https://doi.org/10.1161/STROKEAHA.109.564633.

[29] N. McCormick, V. Bhole, D. Lacaille, J.A. Avina-Zubieta, Validity of diagnostic codes for acute stroke in administrative databases: a systematic review, PLoS One 10 (8) (2015) e0135834, https://doi.org/10.1371/journal.pone.0135834.