SLEEPJ, 2021, 1–13

doi: 10.1093/sleep/zsaa176
Advance Access Publication Date: 14 September 2020
Original Article

ORIGINAL ARTICLE

Small sleepers, big data: leveraging big data to explore sleep-disordered breathing in infants and young children

Zarmina Ehsan[1,2,*], Earl F. Glynn[3], Mark A. Hoffman[2,3], David G. Ingram[1,2] and Baha Al-Shawwa[1,2]

[1]Division of Pulmonary and Sleep Medicine, Children's Mercy-Kansas City, Kansas City, MO, [2]Department of Pediatrics, University of Missouri-Kansas City School of Medicine, Kansas City, MO and [3]Research Informatics, Children's Mercy Research Institute, Children's Mercy-Kansas City, Kansas City, MO

*Corresponding author. Zarmina Ehsan, Pulmonary and Sleep Medicine, Children's Mercy Hospital, 2401 Gillham Road, Kansas City, MO 64108. Email: zehsan@cmh.edu.

Abstract

Study Objectives: Infants represent an understudied minority in sleep-disordered breathing (SDB) research and yet the disease can have a significant impact on health over the formative years of neurocognitive development that follow. Herein we report data on SDB in this population using a big data approach.

Methods: Data were abstracted using the Cerner Health Facts database. Demographics, sleep diagnoses, comorbid medication conditions, healthcare utilization, and economic outcomes are reported.

Results: In a cohort of 68.7 million unique patients, over a 9-year period, there were 9,773 infants and young children with a diagnosis of SDB (obstructive sleep apnea [OSA], nonobstructive sleep apnea, and "other" sleep apnea) who met inclusion criteria, encompassing 17,574 encounters, and a total of 27,290 diagnoses across 62 U.S. health systems, 172 facilities, and 3 patient encounter types (inpatient, clinic, and outpatient). Thirty-nine percent were female. Thirty-nine percent were ≤1 year of age (6,429 infants), 50% were 1–2 years of age, and 11% were 2 years of age. The most common comorbid diagnoses were micrognathia, congenital airway abnormalities, gastroesophageal reflux, chronic tonsillitis/adenoiditis, and anomalies of the respiratory system. Payor mix was dominated by government-funded entities.

Conclusions: We have used a novel resource, large-scale aggregate, de-identified EHR data, to examine SDB. In this population, SDB is multifactorial, closely linked to comorbid medical conditions and may contribute to a significant burden of healthcare costs. Further research focusing on infants at highest risk for SDB can help target resources and facilitate personalized management.

Statement of Significance

There is limited research on sleep-disordered breathing (SDB) in infants and young children. We have used big data methodology to determine clinical demographics and comorbidities for this under studied cohort. Our findings replicate what is known about SDB in this population. Further research in infants and young children at highest risk for disease will be helpful in better understanding long-term outcomes of SDB in this age group and develop treatment guidelines.

Key words: OSA; sleep-disordered breathing; infants; data science; Health Facts

Submitted: 3 February, 2020; **Revised:** 6 August, 2020

Introduction

Although sleep-related breathing disorders are frequently encountered in infants seen in clinical practice, information regarding the prevalence of obstructive sleep apnea (OSA) in infants is limited. Snoring is present in 11.8%–26% of infants, but less than 10% of snoring infants have OSA [1, 2]. Untreated OSA is associated with neurodevelopmental and cognitive dysfunction in older children; however, the long-term outcomes of infant sleep-related breathing disorders are largely unknown [3–5]. The etiology of sleep-disordered breathing (SDB) in infants is multifactorial and pathophysiology is unique from older children and adults [6]. Standardized diagnostic and treatment algorithms are lacking, a hinderance to care for these infants. Clinical examination alone or questionnaires are not reliable in predicting OSA [7, 8]. In infants specifically, there is significant overlap in symptoms of SDB in those with and without OSA. Moreover, children under 2 years of age are at a higher risk of upper airway obstruction (contributing to OSA) and are often symptomatic while awake [9]. Polysomnography (PSG) is the gold standard test to evaluate for OSA but is underutilized in infants due to need for specialty centers and support. Additionally, there is very limited literature regarding PSG data in normal healthy infants especially in young neonates. This lack of data makes it challenging to interpret and quantify abnormal PSG reports. The gaps in currently available literature are (1) limited understanding of the evolution of SDB during infancy and beyond, (2) paucity of centers performing infant PSGs and therefore available literature generated based on information from select centers, and (3) management of infant OSA based on extrapolation of available literature from older children, potentially increasing the risk of providing inappropriate therapy to this understudied and underserved population with SDB. These gaps also make this population an ideal focus of data science. There has been increasing interest in the interplay of big data methodology within sleep medicine [10]. To this effect, the National Sleep Research Resource was developed as a comprehensive data sharing warehouse to facilitate sleep research [11]. However, it only includes NIH-funded prospective studies, of which only three include children, and none of which include those under 2 years of age. Electronic health records are near-universally used across the United States and have a pivotal role within sleep research. Cerner Health Facts (HF) is a de-identified electronic health record system encompassing data from over 600 participating Cerner client facilities within the United States, representing over 68 million unique patients. There is no literature reporting the use of HF for sleep research. The goals of this study were to determine the association between SDB in infants and young children and comorbid medical disorders, characterize the patterns of healthcare utilization, and explore economic outcomes in this population to better understand these relationships and determine priorities for future research focusing on precision medicine in infant SDB. We hypothesized that big data methodology is an effective tool to determine clinical comorbidities associated with infant OSA.

Methods

Data were abstracted using the HF database (Cerner Corporation, Kansas City, MO). Data available in HF include information on patient demographics, diagnoses, medication orders, procedures, and laboratory tests, and details about the type of treatment setting. The institutional review board at Children's Mercy-Kansas City has deemed research with HF to be nonhuman subjects' research.

Three SDB diagnosis cohorts were defined based on International Statistical Classification of Diseases and Related Health Problems (ICD) 9/10 diagnosis codes defined as phenotypes in the online PheWAS mappings (Table 1) using previously described methodology [12, 13]. Health Facts records demographic data by encounter, and therefore patient demographic

Table 1. Diagnosis cohorts defined by sleep apnea phenotypes

Diagnosis cohort	Type	Code	Description
Obstructive sleep apnea (PheCode 327.32)	ICD10-CM	G47.33	Obstructive sleep apnea (adult) (pediatric)
	ICD9	327.23	Obstructive sleep apnea (adult) (pediatric)
Nonobstructive sleep apnea (PheCode 327.31)	ICD10-CM	G47.31	Primary central sleep apnea
	ICD10-CM	G47.34	Idiopathic sleep-related nonobstructive alveolar hypoventilation
	ICD10-CM	G47.36	Sleep-related hypoventilation in conditions classified elsewhere
	ICD10-CM	G47.37	Central sleep apnea in conditions classified elsewhere
	ICD9	327.21	Primary central sleep apnea
	ICD9	327.24	Idiopathic sleep-related nonobstructive alveolar hypoventilation
	ICD9	327.26	Sleep-related hypoventilation/hypoxemia in conditions classifiable elsewhere
	ICD9	327.27	Central sleep apnea in conditions classified elsewhere
(Other) sleep apnea (PheCode 327.3)	ICD10-CM	G47.3	Sleep apnea
	ICD10-CM	G47.30	Sleep apnea, unspecified
	ICD10-CM	G47.39	Other sleep apnea
	ICD9	327.2	Organic sleep apnea
	ICD9	327.20	Organic sleep apnea, unspecified
	ICD9	327.29	Other organic sleep apnea
	ICD9	780.51	Insomnia with sleep apnea, unspecified
	ICD9	780.53	Hypersomnia with sleep apnea, unspecified
	ICD9	780.57	Unspecified sleep apnea

Diagnosis codes for the cohorts are based on International Statistical Classification of Diseases and Related Health Problems (ICD) 9/10 diagnosis codes defined as phenotypes in the online Phenome-Wide Association Studies (PheWAS) database as PheCodes (https://phewascatalog.org/phecodes and https://phewascatalog.org/phecodes_icd10cm). Note, not all ICD 9/10 codes in Health Facts have mappings to PheCodes.

information can vary between encounters. A patient's first encounter with a sleep apnea diagnosis was treated as an index visit to simplify reporting on a patient basis. A separate cohort based on Current Procedural Terminology, 4th Edition (CPT4) sleep procedures involving PSG was also queried (Table 2).

HF data were extracted for the combined diagnosis-based cohorts using a process outlined in Figure 1. Inclusion criteria were: age from 0 to 24 months (with ages 0–12 months defined as infants and those aged >12 to ≤24 months as young children), cared for between the years 2009 and 2017 in either the inpatient, outpatient, or clinic setting (defined as patient encounter types for this manuscript). We focused on this age group for multiple reasons. SDB in children under 2 years of age is typically managed differently from older children [9]. This group is at particular risk of respiratory complications from SDB as well as adverse events from OSA surgery such as adenotonsillectomy. Additionally, management options are not a straightforward and limited guidance exists in the literature. PSG procedure codes were also queried. Of note, the inclusion of procedure codes has only been present within HF since 2013; therefore, procedure data were limited to the years 2013–2017. Encounters were restricted to those with a diagnosis priority ≤ 5 (sleep apnea diagnosis to be present within the top 5 diagnoses for that encounter), which excluded about 17% of the encounters with complex conditions, reducing "noise" in the data. Priorities ranged from 1 to 108 for sleep apnea diagnoses (in HF there can be multiple diagnoses with the same priority).

Table 2. Sleep procedures (polysomnography) as defined by CPT4 codes

CPT4 code	Description
95782	Polysomnography: younger than 6 years, sleep staging with four or more additional parameters of sleep, attended by a technologist
95807	Sleep study, simultaneous recording of ventilation, respiratory effort, ECG or heart rate, and oxygen saturation, attended by a technologist
95808	Polysomnography: sleep staging with 1–3 additional parameters of sleep, attended by a technologist
95810	Polysomnography: sleep staging with four or more additional parameters of sleep, attended by a technologist

CPT4: Current Procedural Terminology, 4th Edition code for polysomnography.

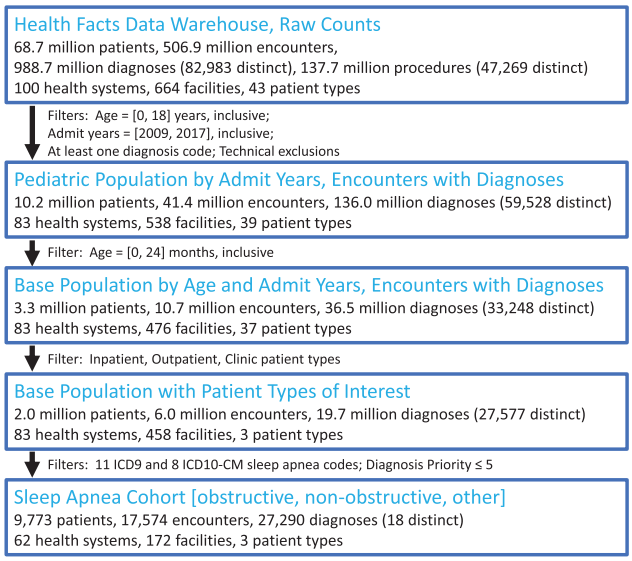


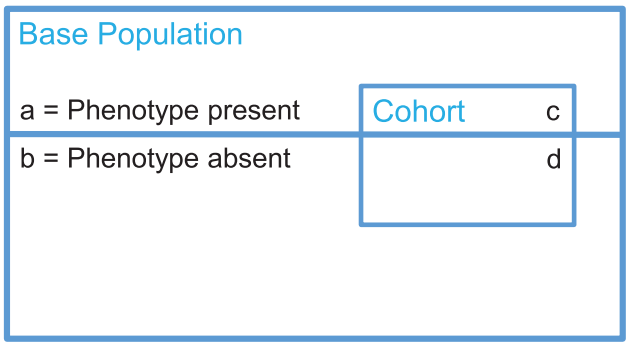
Figure 1. Data flow for selection of sleep apnea cohort from Cerner Health Facts database. This figure summarizes the HF database dataflow used for deriving the sleep apnea cohort. Inclusion and exclusion criteria have been labeled as "filters." Technical exclusions: encounters mapped to by multiple unique patients. Diagnosis priority ≤ 5: sleep apnea diagnosis to be present within the top 5 diagnoses for that encounter. ICD9 and ICD10 codes have been presented in Table 1 of the manuscript.

The combined cohort was split into individual SDB cohorts in the study of comorbidities. Cohort demographics were analyzed, including interrelationships among the cohorts. The diagnosis codes for all encounters in the base population (inclusive of all children within our age group of interest but without sleep apnea diagnoses) from Figure 1 were extracted as the basis for the comorbidity analysis. We used the phenotype codes defined by Denny et al. [12] and Wu et al. [13] simplify the ICD code analysis by collapsing the over 27,000 diagnosis codes into less than 2,000 phenotype codes (which generally aggregate to body systems and organs). The use of phenotype codes provides a good first approach to aggregate comorbidities when no other information is available [14, 15]. We used UpSet plots [16] to depict our results. Multidimensional Venn diagrams can be difficult to draw and interpret. UpSet plots provide a better way to display and interpret set intersections when working with multiple dimensions. The horizontal bars in an UpSet plot are the initial sets, like the circles in a Venn diagram. The vertical bars show the unique combinations of contributions from the original sets (the overlapping areas in a Venn diagram), where the unique combinations are indicated by the filled-in circles. The sum of the counts from the vertical bars gives the number from all the sets.

A ratio of cohort prevalence to base population prevalence revealed comorbidities for phenotypes associated with each sleep apnea cohort. We used a version of Health Facts hosted on Microsoft (Redmond, WA) Azure. Analysis processing was performed using R Studio version 1.2.5033 with R version 3.6.0 [17,18].

The prevalence ratio (Figure 2) is a measure of the strength of the association between the cohort and base population for a given phenotype. This ratio will be 1.0 if the prevalence is the same in both populations. The ratio will be greater than 1.0 when the cohort prevalence is greater than base population prevalence, or less than 1.0 when the cohort prevalence is lower than the base population [19]. These may be easier to interpret than odds ratios [20]. Prevalence ratio confidence intervals were computed from 1,000 nonparametric bootstrap samples of the 6 million encounters in the base population for each of the sleep apnea cohorts (obstructive, nonobstructive, other) [21]. The encounters only with the specific cohort of interest were identified in each bootstrap sample. Prevalence ratios were computed for each phenotype present as shown by the formula in Figure 2. A phenotype's 95% confidence interval was computed from the range of the 2.5% and 97.5% quantiles of the 1,000 bootstrap prevalence ratios.

For comorbidity analysis each of the sleep apnea cohorts was adjusted to include only specific diagnoses. That is, encounters in the adjusted OSA cohort were excluded if they also reported nonobstructive or other sleep apnea. The goal was to make the adjusted cohort "pure" with only the diagnoses of interest. With multiple encounters per patient it was possible to have the other cohort diagnoses as comorbidities.



$$Prevalence\ Ratio = \frac{Cohort\ Prevalence}{Base\ Prevalence} = \frac{\frac{c}{c+d}}{\frac{a}{a+b}}$$

Figure 2. Calculation of the prevalence ratio for comorbidities. The cohort is a subset of the base population and contains all the patients with the phenotype (comorbidity) of interest. The base population is all children ≤ 2 years (~2 million). Definitions—*a*: patient count with phenotype (comorbidity) present in base population; *b*: patient count with phenotype (comorbidity) absent in base population; *c*: patient count with comorbidity present in cohort patients; *d*: patient count with comorbidity absent in cohort patients.

Results

Patient demographics

In a base population of approximately 2 million infants and young children, 9,773 patients across 172 facilities were included. In the 9,773 index patients, 3,807 (39.0%) were female, 5,962 (61.0%) were male. The gender for four patients is not known.

Thirty-nine percent were ≤1 year of age (6,429 infants), 50% were 1–2 years of age, and 11% were 2 years of age. The race breakdown was White (Cerner uses the term Caucasian) (4,474, 46%), African American (2,361, 24%), Hispanic (233, 2.4%), Biracial (181, 1.9%), Asian (94, 1.0%), Pacific Islander (90, 0.9%), and Native American (62, 0.6%). The race for the remaining 2,278 (23.3%) are not known. The breakdown of index patients by U.S. Census region was as follows: Midwest (3,197, 32.7%), South (3,164, 32.4%), West (2,101, 21.5%), and Northeast (1,311, 13.4%).

Cohorts

Of the 9,773 patients included in the study, 7,328 had a diagnosis of OSA, 3,337 had "other" sleep apnea, and 1,164 had nonobstructive sleep apnea (Table 3). Of note, each patient on average has about 1.8 encounters with multiple diagnoses possibly at different facilities, which complicates column totals without separate database queries. Furthermore, these patients belonged to unique combinations of cohorts which are represented in an UpSet plot [16] (Supplementary Figure 1). Seventeen facilities performed the infant/toddler sleep procedures (PSG) and reported the three types of sleep apnea diagnoses (Supplementary Figure 2). The remaining facilities reported various combinations of sleep apnea diagnoses but no procedures. Nine out of 17 facilities reporting sleep procedures were at putative pediatric facilities, defined by a mean encounter age <21 years (Supplementary Figure 3). For HF facilities reporting diagnoses and procedures related to sleep disorder encounters in this study, most reported both diagnoses and procedures (98% of adult facilities and 93% of pediatric facilities), but for the specific sleep disorders in this study 68% of pediatric facilities reported only diagnoses without procedures. For adult facilities, 94% reported only sleep diagnoses without procedures (Table 4).

Patient types

Patients in the cohort were seen across three settings encompassing patient care ("patient-type" designation: inpatient, outpatient, and clinic) (Figure 3). The outpatient group was the largest in each of the cohorts. Only 22 patients were seen for OSA with all three patient-type designations. No patient was reported with all three patient types in the other cohorts.

Comorbidities

Summary of base population comorbidities

The prevalence of OSA, nonobstructive sleep apnea, and "other" sleep apnea in the base population was approximately 0.4%, 0.1%, and 0.2%, respectively.

Cohort comorbidities

The top 30 of 919 comorbidities based on prevalence in the OSA cohort are shown in Table 5, which shows prevalence ratios comparing the prevalence of a phenotype in the cohort to the prevalence in the base population. Similar results for nonobstructive and "other" apnea are shown in Tables 6 and 7. Because OSA comorbidities over infancy may vary according to age, we have also reported comorbidities across subgroups within the 0–24 months age (Table 8).

Prevalence ratio

Table 5 show phenotypes that are 10 times or more prevalent (an order of magnitude) in the OSA cohort when compared to the base population. OSA comorbidities with the highest prevalence ratios in this table include: (1) chronic tonsillitis and adenoiditis (ratio 36.0); (2) anomalies of jaw size/symmetry (ratio 27.9); (3) anomalies of the respiratory system (ratio 14.6); (4) dependence on a respirator (ventilator) (ratio 11.8); (5) other diseases of the lung (ratio 11.6); and (6) chromosomal anomalies. Here is an example of computing prevalence ratio for the chronic tonsillitis and adenoiditis phenotype in Table 5, using the formula from Figure 2.

$$\text{Prevalence ratio} = (1,499/6,950)/(12,281/2,048,911) = 21.57\%/0.60\% = 36.0$$

This phenotype is 36 times more prevalent in the cohort than the base population. Prevalence ratios for nonobstructive sleep apnea and other sleep apnea have been provided in Tables 6 and 7, respectively.

Figure 4 shows that the $\log_2$ prevalence ratios in the OSA cohort are normally distributed (919 total comorbidities, only 30 shown in Table 5). Therefore, with a mean of 0.33 and *SD* of 1.85, comorbidities with $\log_2$ prevalence ratios of 0.33 + 1.85 = 2.18 or more can be considered "high" and clinically correlated for medical significance. In Table 5, we have highlighted comorbidities with a prevalence 10 times that in the general population (corresponding to a $\log_2$ prevalence ratio of 3.3).

Payor mix and total charges

HF reports more than two dozen payor categories (listed below), which were recoded into five: Commercial, Government, Other, Self-Pay, Unknown.

Table 3. Health Facts query result summary for sleep cohort (including procedures)

Type	PheCode	Cohort	Encounters	Patients	Facilities
Diagnosis	327.32	Obstructive sleep apnea	12,758	7,328	136
	327.31	Nonobstructive sleep apnea	2,001	1,164	64
	327.3	(Other) sleep apnea	4,201	3,337	131
		Sleep apnea cohort	**17,574**	**9,773**	**172**
Procedure	N/A	Sleep procedures	812	727	17
All		**Sleep cohort**	**17,672**	**9,840**	**172**

The total number of patients, encounters, and facilities with the three sleep apnea diagnoses and polysomnography (procedures) in our cohorts.

Table 4. Facility differences in reporting sleep apnea diagnoses and procedures in Health Facts

Category	Facilities	Diagnoses only	Diagnoses and procedures	Sleep diagnoses only	Sleep diagnoses and procedures
Sleep facilities	172	2.9%	97.1%	90.1%	9.9%
Sleep adult facilities	144	2.1%	97.9%	94.4%	5.6%
Sleep pediatric facilities	28	7.1%	92.9%	67.9%	32.1%

Distribution of HF facilities reporting sleep apnea diagnoses and polysomnography procedures.

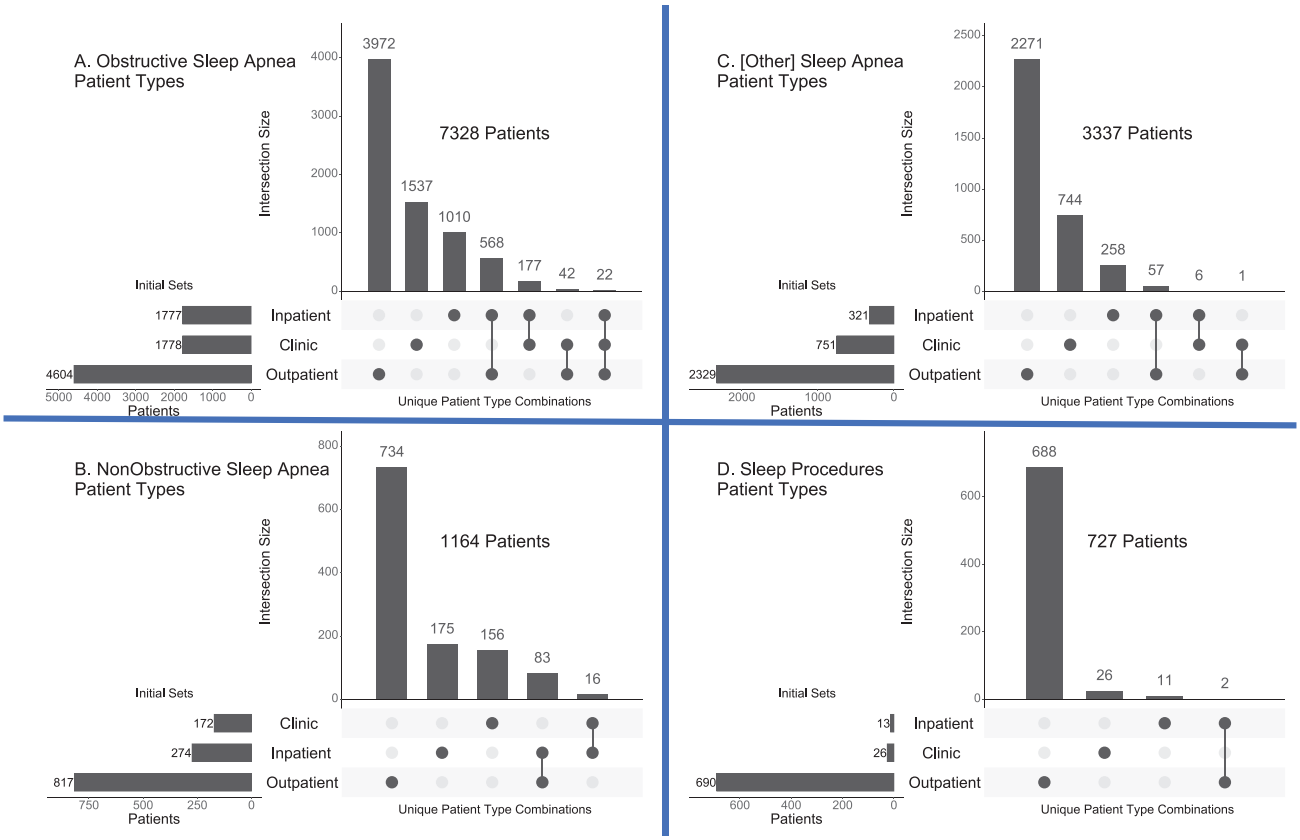


Figure 3. UpSet plots of sleep apnea cohort and sleep procedures. UpSet plot depicting breakdown of patient counts by patient type for each of the three sleep apnea diagnosis cohorts and the sleep procedure cohort. (A) Obstructive sleep apnea; (B) nonobstructive sleep apnea; (C) (Other) sleep apnea; (D) sleep procedures. The horizontal bars reflect patient counts in initial cohorts, which can include duplicate counts for the same patient. The vertical bars show unique counts of patients for the various patient-type combinations. Some facilities used "clinic" as a designation instead of "outpatient." Although these patient types can be combined, sometimes there are differences observed between them.

Payor information in HF is reported with most encounters. In our study 51.4% of "other" sleep apnea encounters received government funding, while 66.1% of nonobstructive apnea encounters did. Self-pay was low ranging from 0.85% for nonobstructive to 1.74% for "other" apnea (Figure 5).

Fewer than 20% of HF encounters report "total charges" (Figure 6). No additional breakdown is available regardless of the length-of-stay or complexity of cases, which explains the large variability in the data. Range values may not be reliable, but median total charges can be useful, especially after filtering for various factors (e.g. length-of-stay, multiple procedures, or diagnoses). Median total charges for sleep disorder encounters for a commercial payor ranged from $47.5K for OSA to $64.6K for nonobstructive sleep apnea. Total

Table 5. "Top 30" obstructive sleep apnea comorbidities

Cohort prevalence rank	PheCode	Phenotype	Cohort patients	Cohort prevalence	Base patients	Base prevalence	Prevalence ratio	95% confidence interval
1	474.2	Chronic tonsillitis and adenoiditis	1,499	21.57%	12,281	0.60%	36.0	36.0, 38.3
2	530.11	GERD	984	14.16%	78,168	3.82%	3.7	3.6, 3.9
3	512.9	Other dyspnea	824	11.86%	31,015	1.51%	7.8	7.8, 8.5
4	748	Anomalies of respiratory system, congenital	814	11.71%	16,406	0.80%	14.6	13.9, 15.1
5	479	Other upper respiratory disease	574	8.26%	47,380	2.31%	3.6	3.5, 3.9
6	495	Asthma	555	7.99%	38,256	1.87%	4.3	4.0, 4.5
7	1013	Asphyxia and hypoxemia	534	7.68%	28,353	1.38%	5.6	5.8, 6.4
8	758.1	Chromosomal anomalies	491	7.06%	12,691	0.62%	11.4	10.7, 11.8
9	1002	Symptoms concerning nutrition, metabolism, and development	488	7.02%	74,188	3.62%	1.9	1.8, 2.0
10	532	Dysphagia	426	6.13%	21,070	1.03%	6.0	5.6, 6.3
11	381.1	Otitis media	422	6.07%	43,900	2.14%	2.8	2.6, 3.0
12	656.2	Respiratory conditions of fetus and newborn	408	5.87%	63,808	3.11%	1.9	2.0, 2.2
13	381.11	Suppurative and unspecified otitis media	372	5.35%	133,256	6.50%	0.8	0.7, 0.8
14	510	Other diseases of lung	358	5.15%	9,096	0.44%	11.6	11.5, 12.9
15	747.11	Cardiac shunt/heart septal defect	331	4.76%	63,130	3.08%	1.5	1.4, 1.6
16	315	Developmental delays and disorders	311	4.47%	22,511	1.10%	4.1	3.6, 4.2
17	264.2	Failure to thrive (childhood)	305	4.39%	44,328	2.16%	2.0	1.9, 2.1
18	381.2	Eustachian tube disorders	265	3.81%	15,084	0.74%	5.2	4.6, 5.4
19	509.1	Respiratory failure	252	3.63%	16,635	0.81%	4.5	4.2, 5.0
20	1010	Other tests	240	3.45%	572,831	27.96%	0.1	0.1, 0.1
21	749.1	Cleft palate	240	3.45%	7,184	0.35%	9.8	8.8, 10.1
22	509.8	Dependence on respirator (ventilator) or supplemental oxygen	230	3.31%	5,741	0.28%	11.8	11.1, 13.0
23	637	Short gestation; low birth weight; and fetal growth retardation	230	3.31%	171,524	8.37%	0.4	0.4, 0.5
24	465	Acute upper respiratory infections of multiple or unspecified sites	228	3.28%	172,551	8.42%	0.4	0.3, 0.4
25	749.2	Congenital anomalies of skull and face bones	213	3.06%	23,290	1.14%	2.7	2.6, 3.1
26	512.8	Cough	201	2.89%	103,859	5.07%	0.6	0.5, 0.6
27	526.3	Anomalies of jaw size/symmetry	201	2.89%	2,127	0.10%	27.9	27.7, 32.4
28	264.9	Lack of normal physiological development, unspecified	189	2.72%	20,600	1.01%	2.7	2.4, 2.9
29	483	Acute bronchitis and bronchiolitis	183	2.63%	110,413	5.39%	0.5	0.4, 0.5
30	079	Viral infection	179	2.58%	80,373	3.92%	0.7	0.6, 0.7

Phenotype comparisons between cohort and base populations are depicted using a prevalence ratio. Larger values of prevalence ratio for a phenotype suggest higher association with the obstructive sleep apnea cohort. Highlighted prevalence ratios show phenotypes that are 10 times or more prevalent in the cohort than the base population. Here the cohort has been reduced to patients reporting only obstructive sleep apnea diagnosis in an encounter, and not combined with the other sleep apnea diagnoses in the same encounter. "Unknown" comorbidities (ICD 9/10 comorbidities not mapped to phenotype codes) were excluded. For example, the prevalence ratio for chronic tonsillitis and adenoiditis phenotype, is 36 times more prevalent in the cohort than the base population: Prevalence ratio = (1,499/6,950)/(12,281/2,048,911) = 21.568%/0.599% = 36.0.

charges paid by government were more variable than other payor types.

Classification of Health Fact payors:

- Commercial (Blue Cross/Blue Shield, Commercial/Indemnity Insurance, Other Commercial Payer)
- Government (CHAMPUS, Government, Medicaid, Medicated Managed Care, Medicare Psychiatric, Medicare Rehab. Other Government, Title V)
- Other (Free Research, HMO/Managed Cared, International Patient, Managed Care, MIA, Other, Other Non-Govt PPO, Worker's Compensation)
- Self-Pay (Self-Insured, Self-Pay)
- Unknown (Not Mapped, NULL, Unknown/Missing/Invalid)

Discussion

Using a big data approach on this understudied population within sleep research, we confirmed that SDB in infants and young children is multifactorial and closely linked to comorbid medical conditions. Additionally, this is the first study to report use of big data in this age group and to demonstrate the utility of the large-scale de-identified EHR data in sleep research in general. The use of this resource required the application of data science visualization methods including UpSet graphs to understand complex overlapping groups. Likewise, at the scale of data available through HF, we applied a prevalence ratio method to evaluate the changes in frequencies of comorbidities. This is novel as comorbidities with the highest prevalence ratios were also those that are known risk factors for OSA in this age group. This method has potential for future research as it may account for some of the complex factors, including contributing site heterogeneity, that can influence the data and make it difficult to apply traditional statistical methods [14].

While some of the clinical comorbidities encountered in our data set are not surprising, there are others that are less well-known in the medical literature. Chronic tonsillitis and adenoiditis were prevalent in 21% of our OSA cohort compared

Table 6. "Top 30" nonobstructive sleep apnea comorbidities

Cohort prevalence rank	PheCode	Phenotype	Cohort patients	Cohort prevalence	Base patients	Base prevalence	Prevalence ratio	95% confidence interval
1	530.11	GERD	180	23.747%	78,168	3.815%	6.2	5.7, 6.9
2	656.2	Respiratory conditions of fetus and newborn	133	17.546%	63,808	3.114%	5.6	5.5, 6.8
3	510	Other diseases of lung	132	17.414%	9,096	0.444%	39.2	36.0, 44.7
4	748	Anomalies of respiratory system, congenital	89	11.741%	16,406	0.801%	14.7	12.6, 16.5
5	327.32	Obstructive sleep apnea	87	11.478%	7,962	0.389%	29.5	24.2, 32.3
6	532	Dysphagia	81	10.686%	21,070	1.028%	10.4	8.5, 11.4
7	509.8	Dependence on respirator (ventilator) or supplemental oxygen	76	10.026%	5,741	0.280%	35.8	31.2, 41.4
8	747.11	Cardiac shunt/heart septal defect	73	9.631%	63,130	3.081%	3.1	2.7, 3.6
9	1002	Symptoms concerning nutrition, metabolism, and development	72	9.499%	74,188	3.621%	2.6	2.2, 3.0
10	1013	Asphyxia and hypoxemia	69	9.103%	28,353	1.384%	6.6	5.8, 8.0
11	512.9	Other dyspnea	68	8.971%	31,015	1.514%	5.9	5.3, 7.3
12	758.1	Chromosomal anomalies	65	8.575%	12,691	0.619%	13.8	10.6, 14.4
13	637	Short gestation; low birth weight; and fetal growth retardation	64	8.443%	171,524	8.371%	1.0	0.9, 1.3
14	495	Asthma	61	8.047%	38,256	1.867%	4.3	3.7, 5.1
15	315	Developmental delays and disorders	52	6.860%	22,511	1.099%	6.2	5.0, 7.2
16	513.31	Apnea	50	6.596%	10,247	0.500%	13.2	11.2, 16.4
17	264.2	Failure to thrive (childhood)	48	6.332%	44,328	2.163%	2.9	2.3, 3.4
18	747.13	Congenital anomalies of great vessels	39	5.145%	36,860	1.799%	2.9	2.2, 3.5
19	479	Other upper respiratory disease	37	4.881%	47,380	2.312%	2.1	1.7, 2.7
20	509.1	Respiratory failure	37	4.881%	16,635	0.812%	6.0	4.9, 7.8
21	752.2	Other specified congenital anomalies of nervous system	37	4.881%	14,985	0.731%	6.7	5.3, 8.0
22	362.1	Retinopathy of prematurity	36	4.749%	12,249	0.598%	7.9	5.8, 9.4
23	079	Viral infection	33	4.354%	80,373	3.923%	1.1	0.8, 1.4
24	264.9	Lack of normal physiological development, unspecified	31	4.090%	20,600	1.005%	4.1	2.9, 5.0
25	749.1	Cleft palate	31	4.090%	7,184	0.351%	11.7	8.0, 12.9
26	345	Epilepsy, recurrent seizures, convulsions	29	3.826%	7,554	0.369%	10.4	8.2, 12.9
27	759	Other and unspecified congenital anomalies	27	3.562%	10,156	0.496%	7.2	5.5, 8.9
28	345.3	Convulsions	26	3.430%	28,960	1.413%	2.4	1.7, 3.0
29	749.2	Congenital anomalies of skull and face bones	26	3.430%	23,290	1.137%	3.0	2.2, 3.8
30	350.3	Lack of coordination	25	3.298%	8,431	0.411%	8.0	5.7, 10.0

Phenotype comparisons between cohort and base populations are depicted using a prevalence ratio. Larger values of prevalence ratio for a phenotype suggest higher association with the obstructive sleep apnea cohort. Highlighted prevalence ratios show phenotypes that are 10 times or more prevalent in the cohort than the base population. Here the cohort has been reduced to patients reporting only nonobstructive sleep apnea diagnosis in an encounter, and not combined with the other sleep apnea diagnoses in the same encounter. ICD 9/10 comorbidities not mapped to phenotype codes were excluded.

to 0.6% of the base population within this age group (prevalence ratio of 36%). Adenotonsillar hypertrophy is a well-known risk factor for OSA in older children, although few studies also report that it can be prevalent as early as infancy [22, 23]. About 2.9% of our cohort had a diagnosis of jaw abnormality compared to 0.1% in the general population within this age group. This phenotype had second largest prevalence ratio (27.9) but was not present in the nonobstructive sleep apnea or "other "sleep apnea cohorts, which is not unexpected as micrognathia is a known risk factor for OSA in infants particularly in those with a cleft palate [24]. Congenital anomalies of the respiratory system were prevalent in 11% of the OSA cohort compared to 0.8% of the base population. We know from retrospective clinical research that congenital soft tissue airway abnormalities or bone abnormalities involving the face also predispose to OSA. Laryngomalacia is one of the leading causes of stridor, upper airway obstruction, and therefore OSA, in infants. Other airway abnormalities such as subglottic stenosis or tracheomalacia are often present as well [25–27]. Prevalence ratios for "dependence on respirator/ventilator" and "other diseases of the lung" also ranked high for prevalence ratios of 11% each. This is very interesting and suggests that infants and young children with chronic lung conditions may be predisposed to OSA. It also may reflect treatment strategies, which often include mechanical ventilation in this age group. A diagnosis of otitis media was also prevalent within our cohort. There are a handful of studies suggesting a high prevalence of middle ear effusion and otitis media in infants with OSA (particularly those with Down syndrome) [28, 29].

Chromosomal abnormalities consistently scored high on prevalence ratio in all three cohorts with a higher prevalence

Table 7. "Top 30" other sleep apnea comorbidities

Cohort prevalence rank	PheCode	Phenotype	Cohort patients	Cohort prevalence	Base patients	Base prevalence	Prevalence ratio	95% confidence interval
1	474.2	Chronic tonsillitis and adenoiditis	534	17.938%	12,281	0.599%	29.9	31.0, 34.7
2	512.9	Other dyspnea	408	13.705%	31,015	1.514%	9.1	9.3, 10.6
3	530.11	GERD	295	9.909%	78,168	3.815%	2.6	2.5, 2.9
4	381.11	Suppurative and unspecified otitis media	174	5.845%	133,256	6.504%	0.9	0.8, 1.0
5	748	Anomalies of respiratory system, congenital	168	5.643%	16,406	0.801%	7.0	6.4, 7.9
6	495	Asthma	161	5.408%	38,256	1.867%	2.9	2.7, 3.3
7	479	Other upper respiratory disease	154	5.173%	47,380	2.312%	2.2	2.2, 2.7
8	381.1	Otitis media	141	4.736%	43,900	2.143%	2.2	2.0, 2.6
9	532	Dysphagia	130	4.367%	21,070	1.028%	4.2	3.8, 4.8
10	758.1	Chromosomal anomalies	102	3.426%	12,691	0.619%	5.5	4.6, 6.0
11	1002	Symptoms concerning nutrition, metabolism, and development	100	3.359%	74,188	3.621%	0.9	0.8, 1.1
12	327	Sleep disorders	77	2.586%	4,100	0.200%	12.9	11.8, 16.3
13	381.2	Eustachian tube disorders	75	2.519%	15,084	0.736%	3.4	2.9, 4.1
14	293	Symptoms involving head and neck	70	2.351%	4,621	0.226%	10.4	10.0, 13.7
15	1013	Asphyxia and hypoxemia	69	2.318%	28,353	1.384%	1.7	1.6, 2.2
16	315	Developmental delays and disorders	69	2.318%	22,511	1.099%	2.1	1.8, 2.5
17	637	Short gestation; low birth weight; and fetal growth retardation	68	2.284%	171,524	8.371%	0.3	0.3, 0.4
18	510	Other diseases of lung	67	2.251%	9,096	0.444%	5.1	4.8, 6.5
19	512.8	Cough	66	2.217%	103,859	5.069%	0.4	0.4, 0.5
20	656.2	Respiratory conditions of fetus and newborn	65	2.183%	63,808	3.114%	0.7	0.7, 1.0
21	1010	Other tests	64	2.150%	572,831	27.958%	0.1	0.1, 0.1
22	264.2	Failure to thrive (childhood)	64	2.150%	44,328	2.163%	1.0	0.8, 1.2
23	389	Hearing loss	61	2.049%	22,388	1.093%	1.9	1.6, 2.3
24	264.9	Lack of normal physiological development, unspecified	58	1.948%	20,600	1.005%	1.9	1.6, 2.3
25	389.2	Conductive hearing loss	57	1.915%	10,739	0.524%	3.7	3.2, 4.7
26	509.8	Dependence on respirator (ventilator) or supplemental oxygen	55	1.847%	5,741	0.280%	6.6	5.5, 8.0
27	315.2	Speech and language disorder	54	1.814%	26,379	1.287%	1.4	1.2, 1.8
28	513.31	Apnea	52	1.747%	10,247	0.500%	3.5	3.2, 4.7
29	465	Acute upper respiratory infections of multiple or unspecified sites	48	1.612%	172,551	8.422%	0.2	0.1, 0.2
30	519.8	Other diseases of respiratory system, NEC	48	1.612%	11,514	0.562%	2.9	2.5, 3.8

Phenotype comparisons between cohort and base populations are depicted using a prevalence ratio. Larger values of prevalence ratio for a phenotype suggest higher association with the obstructive sleep apnea cohort. Highlighted prevalence ratios show phenotypes that are 10 times or more prevalent in the cohort than the base population. Here the cohort has been reduced to patients reporting only "other sleep apnea" diagnosis in an encounter, and not combined with the other sleep apnea diagnoses in the same encounter. ICD 9/10 comorbidities not mapped to phenotype codes were excluded.

than that in the base population. We know from current literature that up to 70% of children with craniosynostosis syndromes (such as Apert syndrome, Crouzon syndrome, Pfeiffer syndrome) have OSA [30]. Similarly, those with craniofacial syndromes and other congenital abnormalities such as achondroplasia, Down syndrome, Beckwith Weidman syndrome, Prader-Willi syndrome are at risk of OSA during infancy and beyond [22].

The link between gastroesophageal reflux and OSA in infants has been reported in the literature; however, the evidence for this is weak [25]. In all three cohorts, we found a higher prevalence of gastroesophageal reflux compared to the base population; however, prevalence ratios did not appear to be significant.

Nonobstructive sleep apnea more commonly represents central sleep apnea. Central sleep apnea is a sleep-related disorder that occurs when there is reduced or absent respiratory effort during sleep. In children, it is mostly associated with underlying medical comorbidities. For those under 2 years of age, the most common etiologies are prematurity and neurologic conditions [31]. Concomitant OSA, dependence on a ventilator and anomalies of the respiratory system ranked highest in prevalence ratios highlighting the overlap between obstructive and central sleep apnea that is often seen in infants as well as the interplay of chronic pulmonary disease and control of breathing disorders in this age group. A few comorbid diagnoses were unique to the nonobstructive sleep apnea cohort. Epilepsy showed a higher prevalence in the nonobstructive sleep apnea cohort compared to the OSA cohort and the base population. We know from traditional research and clinical data that neurologic abnormalities are frequently seen in infants and young children with central sleep apnea [31].

Retinopathy of prematurity was prevalent in about 5% of those with nonobstructive sleep apnea compared to 0.6% of the

Table 8. "Top 10" obstructive sleep apnea comorbidities according to age based on prevalence with comparisons to base population

Cohort prevalence rank	PheCode	Phenotype	Cohort patients	Cohort prevalence	Base patients	Base prevalence	Prevalence ratio	95% confidence interval	
Age 0 to <1 month									
1	656.3	Endocrine and metabolic disturbances of fetus and newborn	78	35.94%	1,48,820	12.08%	3	2.6	3.5
2	656.2	Respiratory conditions of fetus and newborn	64	29.49%	55,807	4.53%	6.5	5.8	8.2
3	747.11	Cardiac shunt/heart septal defect	63	29.03%	33,385	2.71%	10.7	9.6	13.1
4	748	Anomalies of respiratory system, congenital	60	27.65%	4,141	0.34%	82.2	73.7	102.4
5	656	Other perinatal conditions of fetus or newborn	51	23.50%	1,03,505	8.40%	2.8	2.3	3.5
6	526.3	Anomalies of jaw size/symmetry	50	23.04%	1,294	0.11%	219.3	190.1	273.1
7	637	Short gestation; low birth weight; and fetal growth retardation	47	21.66%	1,54,764	12.57%	1.7	1.5	2.2
8	656.26	Transitory tachypnea or apnea of newborn	40	18.43%	54,207	4.40%	4.2	3.5	5.5
9	747.13	Congenital anomalies of great vessels	40	18.43%	23,753	1.93%	9.6	7.9	12.3
10	1002	Symptoms concerning nutrition, metabolism, and development	39	17.97%	31,354	2.55%	7.1	5.5	8.5
Age ≥ 1 to <6 months									
1	748	Anomalies of respiratory system, congenital	246	22.55%	8,137	1.49%	15.1	15	17.4
2	530.11	GERD	227	20.81%	46,910	8.58%	2.4	2.4	2.8
3	1002	Symptoms concerning nutrition, metabolism, and development	149	13.66%	22,832	4.18%	3.3	3.1	3.9
4	1013	Asphyxia and hypoxemia	126	11.55%	9,254	1.69%	6.8	6.9	8.8
5	479	Other upper respiratory disease	124	11.37%	16,620	3.04%	3.7	3.6	4.6
6	747.11	Cardiac shunt/heart septal defect	99	9.07%	25,715	4.70%	1.9	1.7	2.3
7	264.2	Failure to thrive (childhood)	88	8.07%	15,610	2.86%	2.8	2.6	3.4
8	526.3	Anomalies of jaw size/symmetry	88	8.07%	643	0.12%	68.6	66.2	85.7
9	758.1	Chromosomal anomalies	83	7.61%	4,719	0.86%	8.8	7.5	9.9
10	512.9	Other dyspnea	78	7.15%	8,818	1.61%	4.4	4.1	5.7
Age ≥ 6 to <12 months									
1	530.11	GERD	287	17.23%	21,173	4.33%	4	3.9	4.5
2	748	Anomalies of respiratory system, congenital	234	14.05%	4,976	1.02%	13.8	13.4	15.8
3	474.2	Chronic tonsillitis and adenoiditis	158	9.48%	1,698	0.35%	27.3	27.1	33.1
4	1013	Asphyxia and hypoxemia	157	9.42%	7,111	1.45%	6.5	6.9	8.2
5	512.9	Other dyspnea	154	9.24%	6,430	1.31%	7	6.8	8.6
6	758.1	Chromosomal anomalies	150	9.00%	4,688	0.96%	9.4	8.8	10.7
7	479	Other upper respiratory disease	133	7.98%	12,263	2.51%	3.2	3.1	3.9
8	1002	Symptoms concerning nutrition, metabolism, and development	126	7.56%	13,732	2.81%	2.7	2.5	3.1
9	656.2	Respiratory conditions of fetus and newborn	111	6.66%	4,395	0.90%	7.4	6.6	8.3
10	510	Other diseases of lung	96	5.76%	2,998	0.61%	9.4	9	11.6
Age ≥ 12 to 24 months									
1	474.2	Chronic tonsillitis and adenoiditis	1332	26.56%	10,443	1.44%	18.5	19.5	20.8
2	512.9	Other dyspnea	601	11.98%	12,787	1.76%	6.8	7	7.8
3	530.11	GERD	560	11.17%	16,044	2.21%	5	5	5.5
4	495	Asthma	472	9.41%	28,310	3.90%	2.4	2.4	2.7
5	748	Anomalies of respiratory system, congenital	383	7.64%	4,185	0.58%	13.2	12.6	14.3
6	381.1	Otitis media	363	7.24%	30,581	4.22%	1.7	1.7	1.9
7	758.1	Chromosomal anomalies	352	7.02%	6,584	0.91%	7.7	7.4	8.4
8	479	Other upper respiratory disease	319	6.36%	17,232	2.38%	2.7	2.7	3.2
9	381.11	Suppurative and unspecified otitis media	315	6.28%	91,159	12.57%	0.5	0.5	0.5
10	1013	Asphyxia and hypoxemia	295	5.88%	10,429	1.44%	4.1	4.3	5

Phenotype comparisons between cohort and base populations are depicted using a prevalence ratio. Larger values of prevalence ratio for a phenotype suggest higher association with the obstructive sleep apnea cohort. Highlighted prevalence ratios show phenotypes that are 10 times or more prevalent in the cohort than the base population. Here the cohort has been reduced to patients reporting only obstructive sleep apnea diagnosis in an encounter, and not combined with the other sleep apnea diagnoses in the same encounter.

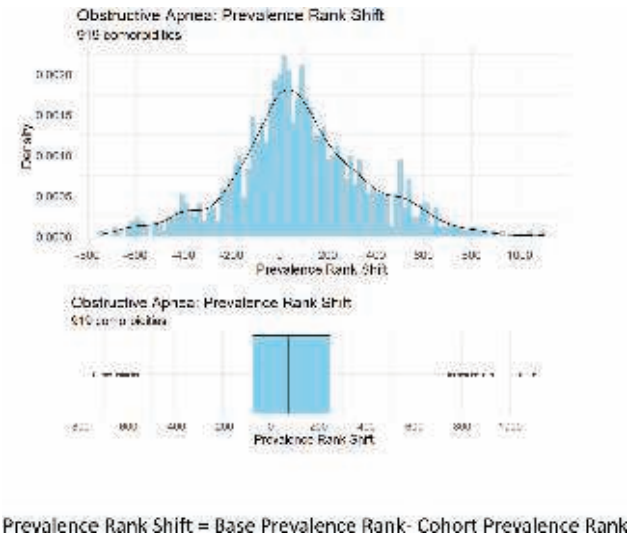


Figure 4. Distribution of $\log_2$ prevalence ratios for obstructive sleep apnea cohort. The prevalence ratio shows how much more common a diagnosis is in a cohort than the base population. The 919 phenotype comorbidities have a log-normal distribution for prevalence ratios as shown by the density plot/histogram (top). The scale is the base-2 logarithm of the prevalence ratio (e.g. a ratio of 1 would have a log ratio of 0, a ratio of 2 would have a log ratio of 1). The corresponding boxplot (bottom) can be used to identify phenotypes with prevalence ratios that are extremely high and could be considered statistical outliers. Phenotypes with high values of prevalence ratios should be evaluated for medical significance.

base population. Prematurity has been cited as a risk factor for OSA in both young and older children and was not found as a top comorbidity in either of our cohorts, but retinopathy of prematurity may perhaps be a surrogate for extreme prematurity in this context [32]. Alternatively, apnea of prematurity or central sleep apnea is highly prevalent in premature infants who are also at risk for retinopathy of prematurity.

The prevalence of failure to thrive was slightly higher in the OSA and nonobstructive sleep apnea cohorts when compared to the base population; however, the prevalence ratio was not significant. Failure to thrive has also been reported in children with OSA and may in fact be the presenting complaint in younger children [33]. Treatment for OSA has also been linked to improvement in growth parameters [33].

In line with available literature, the diagnosis of dysphagia consistently scored high on the prevalence ratio in all three sleep apnea cohorts. Feeding difficulties or dysphagia is frequently present in those with airway obstruction (particularly that which leads to symptoms during both wake and sleep periods) and often improves with treatment of OSA. Infants with adenotonsillar enlargement, laryngomalacia neurologic abnormalities, and micrognathia are particularly at risk [23, 34].

Interestingly, we found a higher prevalence of a diagnosis of "developmental delays and disorders" in the OSA cohort (4.5% versus 1.1%), nonobstructive sleep apnea cohort (6.8% versus 1.1%), and other sleep apnea cohort (2.3% versus 1.1%) compared to the base population. This association may be bidirectional. Children with developmental delay may have generalized hypotonia making their upper airway more vulnerable to collapse. This could be predisposing them to OSA. Additionally, they may have central neurologic abnormalities increasing their risk of nonobstructive or central sleep apnea. Additionally, available

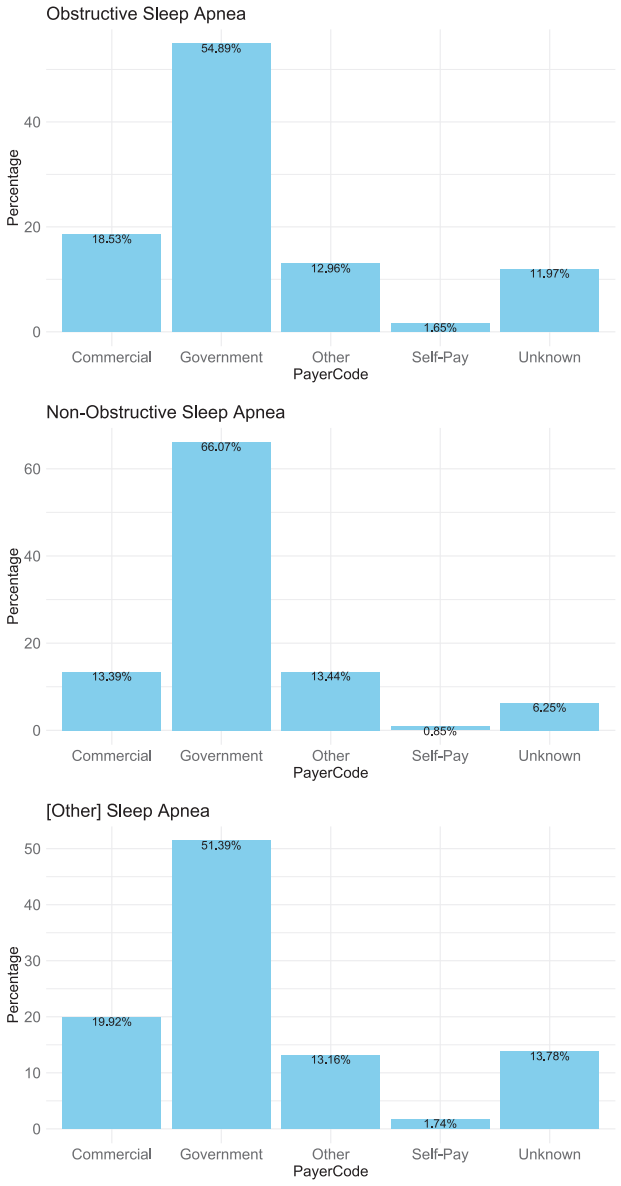


Figure 5. Bar graph depicting payor mix for sleep apnea cohort. Health Facts reports payor information for most encounters broken into more than two dozen categories. The bar graph show sleep apnea cohort summarized into five categories: Commercial, Government, Other, Self-Pay, and Unknown.

literature suggests that OSA in infancy can predispose to behavioral and cognitive delays later in life although the evidence for this is very limited [35–38].

About 9% of infants and young children in the HF data set who had a diagnosis of nonobstructive sleep apnea had diagnoses of a cardiac shunt/heart septal defect compared to 3% of the base population (OSA cohort showed a lower prevalence of 4.7% versus 3% in the base population). Although adults with congenital heart disease are known to have a higher prevalence of sleep apnea, there is only one study in infants suggesting such a link [23, 39]. About 8% of infants and young children with a diagnosis of OSA also had a diagnosis of asthma. While asthma is a known comorbidity in older children with OSA there is no data to suggest such in association in young children and infants [40].

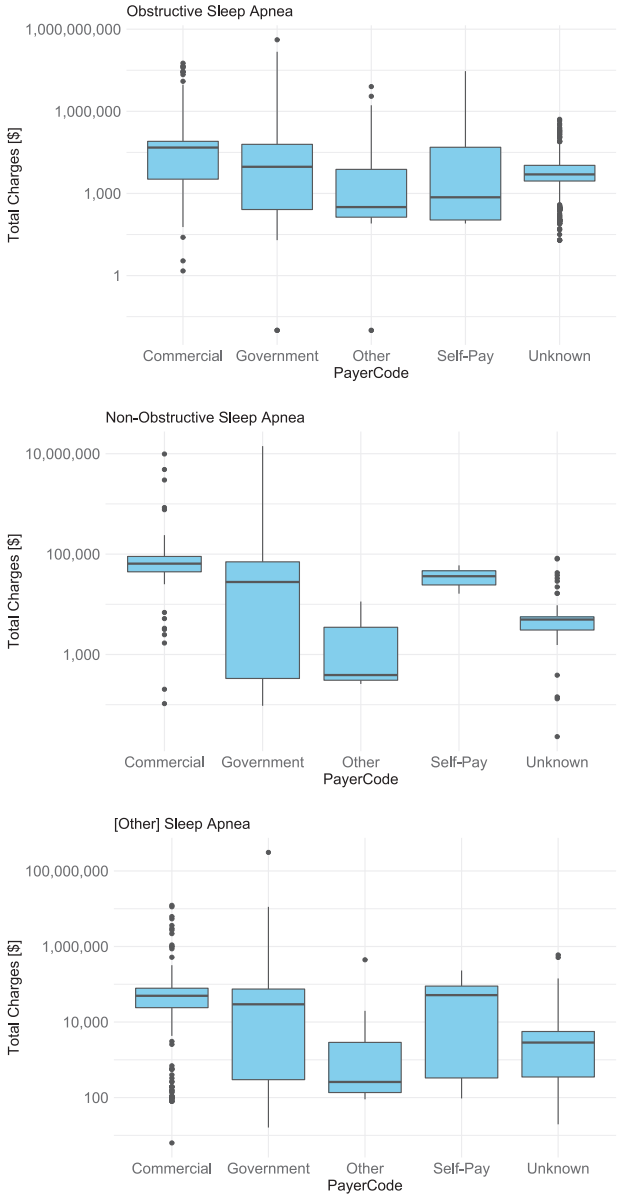


Figure 6. Encounter total charges for sleep apnea cohort. Fewer than 20% of Health Facts encounters report "total charges." No additional breakdown of charges is available regardless of the length-of-stay or complexity of cases, which explains the large variability in the data. Range values may not be reliable, but median total charges may be more robust measures, especially after filtering for various factors (e.g. length-of-stay, multiple procedures, or diagnoses).

Although OSA in adults is associated with a significant increase in economic burden and healthcare utilization, the literature in children is very limited [41, 42]. As we have shown, data regarding healthcare costs of OSA are available in HF. Future studies (within HF) utilizing diagnosis and comorbidity trajectories prior to and after and OSA diagnosis may be helpful in expanding on currently existing literature.

This study has several strengths. We have leveraged data science methodology to focus on an understudied and underrepresented age group within pediatric OSA and are the first to present clinical comorbidities derived in this manner. A cohort size of over 9,000 infants and young children is indeed a large sample size for SDB research within this age group. Current literature (mostly from single-center studies) has reported outcomes in a significantly lower number of infants and young children. In general, our results corroborate what is already known based on traditional research methodology about SDB in infants and young children with a few unusual findings as described earlier. We are the first to report a strong clinical correlation of higher prevalence ratios with disease comorbidities based on phenotypes derived from ICD 9 and 10 codes. Additionally, we are the first report the utility of using HF for sleep research and have highlighted its strengths and limitations therein.

Our study has several limitations. One major limitation is the paucity of polysomnogram data in our cohort. HF only started including procedures as of 2013. It is unclear whether the limited number of polysomnograms reported represents gaps in coding for sleep procedures or whether polysomnograms are indeed underutilized in this age group. Typically, sleep studies in infants are more time-consuming than those in older children and require specialty centers of which there are a few within the United States. The inclusion of detailed polysomnogram results within HF would certainly be ideal for sleep research; however, this capability is not currently addressed within the Cerner EHR system. Another limitation is that a patient can be tracked within different facilities at the same health system, but not across different health systems. Therefore, it is possible that a patient diagnosed with sleep apnea at a HF facility could have had a sleep procedure at facility not tracked by HF. Although one of our objectives was to include economic outcomes, we found that only 20% of encounters in HF reported total charges, which may make it difficult to ascertain meaningful conclusions. In a large de-identified database such as HF there is no way of knowing exactly "why" certain data elements are reported or not. There are many differences by facility caused by local policies, data use agreements with Cerner that can change over time, and data mappings that Cerner can change over time. Encounter "total charges" are reported only in about 20%–25% of encounters, and often the data quality is suspect with extremely large or small values. Encounters can last for hundreds of days with only a single "total charges" composite figure for countless unknown items. Nevertheless, we feel this drawback is important to report as HF is widely used across the U.S. health systems. Lastly, sources of bias such as missing data and misclassification could certainly have affected our results. Such deficiencies within EHR systems have been previously published [43, 44]. As such, the magnitude of the problem is likely underestimated in this study as there is a mixed representation of hospitals.

To conclude, SDB in infants is multifactorial and associated with multiple medical comorbidities. Further research targeting infants and young children is needed to better understand long-term outcomes of SDB in this population. Large-scale aggregate, de-identified EHR data provide a rich and untapped resource to examine these conditions.

Supplementary Material

Supplementary material is available at *SLEEP* online.

Acknowledgments

All authors have seen and approve this manuscript.

Funding

This work was funded by the Division of Pulmonary and Sleep Medicine, Children's Mercy-Kansas City, Kansas City, MO.

Conflict of interest statement. The authors have no relevant conflicts of interest to disclose.

References

1. Kahn A, *et al.* Clinical symptoms associated with brief obstructive sleep apnea in normal infants. *Sleep*. 1993;**16**(5):409–413.
2. Montgomery-Downs HE, *et al.* Sleep habits and risk factors for sleep-disordered breathing in infants and young toddlers in Louisville, Kentucky. *Sleep Med*. 2006;**7**(3):211–219.
3. Jackman AR, *et al.* Sleep disordered breathing in early childhood: quality of life for children and families. *Sleep*. 2013;**36**(11):1639–1646.
4. Baldassari CM, *et al.* Pediatric obstructive sleep apnea and quality of life: a meta-analysis. *Otolaryngol Head Neck Surg*. 2008;**138**(3):265–273.
5. Tan HL, *et al.* Obstructive sleep apnea in children: a critical update. *Nat Sci Sleep*. 2013;**5**:109–123.
6. Katz ES, *et al.* Obstructive sleep apnea in infants. *Am J Respir Crit Care Med*. 2012;**185**(8):805–816.
7. Carroll JL, *et al.* Inability of clinical history to distinguish primary snoring from obstructive sleep-apnea syndrome in children. *Chest*. 1995;**108**(3):610–618.
8. Brietzke SE, *et al.* Can history and physical examination reliably diagnose pediatric obstructive sleep apnea/hypopnea syndrome? A systematic review of the literature. *Otolaryngol Head Neck Surg*. 2004;**131**(6):827–832.
9. Kaditis AG, *et al.* ERS statement on obstructive sleep disordered breathing in 1- to 23-month-old children. *Eur Respir J*. 2017;**50**(6):1700985. doi:10.1183/13993003.00985-2017.
10. Budhiraja R, *et al.* The role of big data in the management of sleep-disordered breathing. *Sleep Med Clin*. 2016;**11**(2):241–255.
11. Zhang GQ, *et al.* The National Sleep Research Resource: towards a sleep data commons. *J Am Med Inform Assoc*. 2018;**25**(10):1351–1358.
12. Denny JC, *et al.* Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;**31**(12):1102–1110.
13. Wu P, *et al.* Mapping ICD-10 and ICD-10-CM codes to Phecodes: workflow development and initial evaluation. *JMIR Med Inform*. 2019;**7**(4):e14325.
14. Glynn EF, *et al.* Heterogeneity introduced by EHR system implementation in a de-identified data resource from 100 non-affiliated organizations. *JAMIA Open*. 2019;**2**(4):554–561.
15. Phecode Maps. PheWAS Resources. https://phewascatalog.org/. Accessed January 17, 2020.
16. Lex A, *et al.* UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph*. 2014;**20**(12):1983–1992.
17. RStudio Team. RStudio: Integrated Development for R. Boston, MA: RStudio, Inc.; 2019. http://www.rstudio.com/. Accessed January 2020.
18. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. https://www.R-project.org/. Accessed January 2020.
19. Martin J. Prevalence Ratio in Cross-Sectional Study. CTSPedia: A Knowledge Base for Clinical and Translational Research; 2009. https://www.ctspedia.org/do/view/CTSpedia/PrevalenceRatio. Accessed April 4, 2020.
20. Martinez B. Odds ratio or prevalence ratio? An overview of reported statistical methods and appropriateness of interpretations in cross-sectional studies with dichotomous outcomes in veterinary medicine. *Front Vet Sci*. 2017;**4**:Article 193. https://www.frontiersin.org/articles/10.3389/fvets.2017.00193/full; https://pubmed.ncbi.nlm.nih.gov/29177157/.2017. Accessed April 4, 2020.
21. Puth MT, *et al.* On the variety of methods for calculating confidence intervals by bootstrapping. *J Anim Ecol*. 2015;**84**(4):892–897.
22. Marcus CL, *et al.*; American Academy of Pediatrics. Diagnosis and management of childhood obstructive sleep apnea syndrome. *Pediatrics*. 2012;**130**(3):e714–e755.
23. Harada G, *et al.* Prevalence and risk factors of sleep apnea in adult patients with congenital heart disease. *Cardiol Young*. 2019;**29**(5):576–582.
24. Breugem CC, *et al.* Best practices for the diagnosis and evaluation of infants with robin sequence: a clinical consensus report. *JAMA Pediatr*. 2016;**170**(9):894–902.
25. Qubty WF, *et al.* Comorbidities in infants with obstructive sleep apnea. *J Clin Sleep Med*. 2014;**10**(11):1213–1216.
26. Camacho M, *et al.* Supraglottoplasty for laryngomalacia with obstructive sleep apnea: a systematic review and meta-analysis. *Laryngoscope*. 2016;**126**(5):1246–1255.
27. Farhood Z, *et al.* Objective outcomes of supraglottoplasty for children with laryngomalacia and obstructive sleep apnea: a meta-analysis. *JAMA Otolaryngol Head Neck Surg*. 2016;**142**(7):665–671.
28. Mitchell RB, *et al.* Ear, nose and throat disorders in children with Down syndrome. *Laryngoscope*. 2003;**113**(2):259–263.
29. Robison JG, *et al.* Increased eustachian tube dysfunction in infants with obstructive sleep apnea. *Laryngoscope*. 2012;**122**(5):1170–1177.
30. Driessen C, *et al.* How does obstructive sleep apnoea evolve in syndromic craniosynostosis? A prospective cohort study. *Arch Dis Child*. 2013;**98**(7):538–543.
31. McLaren AT, *et al.* Diagnosis, management and pathophysiology of central sleep apnea in children. *Paediatr Respir Rev*. 2019;**30**:49–57.
32. Raynes-Greenow CH, *et al.* Sleep apnea in early childhood associated with preterm birth but not small for gestational age: a population-based record linkage study. *Sleep*. 2012;**35**(11):1475–1480.
33. Bonuck KA, *et al.* Growth and growth biomarker changes after adenotonsillectomy: systematic review and meta-analysis. *Arch Dis Child*. 2009;**94**(2):83–91.
34. Daniel M, *et al.* Airway, feeding and growth in infants with Robin sequence and sleep apnoea. *Int J Pediatr Otorhinolaryngol*. 2013;**77**(4):499–503.
35. Bonuck K, *et al.* Sleep-disordered breathing in a population-based cohort: behavioral outcomes at 4 and 7 years. *Pediatrics*. 2012;**129**(4):e857–e865.
36. Bonuck KA, *et al.* Prevalence and persistence of sleep disordered breathing symptoms in young children: a 6-year population-based cohort study. *Sleep*. 2011;**34**(7):875–884.
37. Montgomery-Downs HE, *et al.* Snore-associated sleep fragmentation in infancy: mental development effects and

contribution of secondhand cigarette smoke exposure. *Pediatrics*. 2006;**117**(3):e496–e502.
38. Piteo AM, *et al*. Snoring and cognitive development in infancy. *Sleep Med*. 2011;**12**(10):981–987.
39. Combs D, *et al*. Sleep-disordered breathing is associated with increased mortality in hospitalized infants with congenital heart disease. *J Clin Sleep Med*. 2018;**14**(9):1551–1558.
40. Davies SE, *et al*. The association between asthma and obstructive sleep apnea (OSA): a systematic review. *J Asthma*. 2019;**56**(2):118–129.
41. Tarasiuk A, *et al*. Elevated morbidity and health care use in children with obstructive sleep apnea syndrome. *Am J Respir Crit Care Med*. 2007;**175**(1):55–61.
42. Tarasiuk A, *et al*. Adenotonsillectomy in children with obstructive sleep apnea syndrome reduces health care utilization. *Pediatrics*. 2004;**113**(2):351–356.
43. Gianfrancesco MA, *et al*. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med*. 2018;**178**(11):1544–1547.
44. Graber ML, *et al*. The impact of electronic health records on diagnosis. *Diagnosis (Berl)*. 2017;**4**(4):211–223.