

Identifying Psychiatric Comorbidities for Obstructive Sleep Apnea in the Biomedical Literature and Electronic Health Record

Jessica M. Haddad, BS¹ and Elizabeth S. Chen, PhD¹

¹Center for Biomedical Informatics, Brown University, Providence, RI

Abstract

Obstructive sleep apnea (OSA) is one of the most common diseases among Americans, affecting between 5 and 20% of the population. While there is existing evidence of numerous comorbid conditions, such as obesity, diabetes, and high blood pressure, the vast majority of this evidence has focused explicitly on cardiovascular morbidities and excluded any mental or behavioral disorders. The goal of this study was to examine psychiatric comorbidities of OSA in two types of sources: (1) biomedical literature in the MEDLINE/PubMed database (focusing on MeSH descriptors) and Semantic MEDLINE Database (SemMedDB; for semantic predications), and (2) electronic health record data in the MIMIC-III database. Approximately 300 unique psychiatric comorbidities were identified, ranked, and compared across MEDLINE/PubMed, SemMedDB, and MIMIC-III. The preliminary results highlight the potential of this multi-angled approach for suggesting opportunities for further investigation that may contribute to improving mental health in persons afflicted with OSA.

Introduction

Obstructive sleep apnea (OSA) is a sleep-related breathing disorder characterized by repeated partial or complete closing of the airway when patients lie down to sleep, and is highly prevalent in middle-aged overweight patients, especially men.¹ Like most sleep disorders, OSA impairs many aspects of mental and physical health due to the vital role of adequate sleep hygiene. Psychiatric comorbidity in OSA affects adherence to treatment like continuous positive airway pressure (CPAP) as well as patient quality of life.¹

Previous studies have found moderate evidence for associations between OSA and psychiatric comorbidities. For instance, Huang *et al.* mined data from the New South Wales Inpatient Data Collection repository collected between 1999 and 2004, in which clinical diagnoses were coded with ICD-10-CM codes.² Most of the comorbidities they found were purely physiological (e.g., obesity and Type II Diabetes), but they did find depressive episodes to be a common comorbid condition associated with OSA, occurring with 2% frequency (out of 60,197 cases). More broadly, they also found a 3-4% frequency of mental and behavioral disorders, being slightly more common in females than males. The authors explain that mental problems (e.g., anxiety, depression, and psychosis) may arise in OSA patients due to low oxygen saturation levels during apneic events, severe sleep fragmentation, and excessive daytime sleepiness.

Sharafkhaneh, Giray, and Richardson utilized Veterans Health Administration data from 1998 to 2001, containing ICD-9-CM codes for over 4 million cases.³ About 2.91% of the cases had OSA (118,105 individuals). Psychiatric comorbid diagnoses in the sleep apnea group included depression (21.8%), anxiety (16.7%), posttraumatic stress disorder (11.9%), psychosis (5.1%), and bipolar disorders (3.3%). Compared with patients not diagnosed with sleep apnea, a significantly greater prevalence ($P < .0001$) was found for mood disorders, anxiety, posttraumatic stress disorder (PTSD), psychosis, and dementia in patients with sleep apnea.

Finally, Gupta and Simpson searched the PubMed, EMBASE, and PsycINFO databases using ICD-9-CM codes for OSA and DSM-IV-TR diagnostic groups, retrieving 48 relevant records in which a meta-analysis was conducted.⁴ Their findings indicated an increased prevalence of OSA in individuals with major depressive disorder (MDD) and PTSD, although they acknowledge having moderate heterogeneity and a high risk of bias. They found insufficient evidence to support increased OSA in schizophrenia and psychotic disorders, bipolar and related disorders, and anxiety disorders other than PTSD. The authors also summarized how different treatments and interventions have been utilized in patient populations with specific mental disorders.

There is a limited amount of research looking at the effects of OSA on mental health. As such, the goal of this study was to utilize large data sets in order to verify and amend prior research on the subject. In particular, we aim to examine the prevalence of psychiatric comorbidities in OSA populations and hypothesize that there are associations between OSA and several psychiatric disorders, such as depression and anxiety disorders.

Methods

In this study, three biomedical and clinical data sources were analyzed to identify, rank, and compare psychiatric comorbidities for OSA.

Data Sources

1. MEDLINE/PubMed

The MEDLINE citation database is a leading source of biomedical literature and references. MEDLINE, the primary component of PubMed, contains biomedical references including abstracts for articles in thousands of biomedical journals, and holds over 26 million citations.⁵⁻⁶ Each record is composed of a title, abstract, and Medical Subject Headings (MeSH) descriptors in order to index MEDLINE citations. MeSH descriptors are valuable metadata as they function as a window into the full text. Various systems have been developed to identify and visualize relationships between MeSH descriptors in MEDLINE.⁷

2. Semantic MEDLINE Database

The Semantic MEDLINE Database (SemMedDB⁸⁻⁹) uses SemRep¹⁰, a semantic interpreter of biomedical text, to extract semantic predications (subject-predicate-object triples) from titles and abstracts. Currently, the repository contains information for over 82.2 million predications from the 26 million PubMed citations. The database contains an array of imperative information about the citations (e.g., PMID and publication date), UMLS (Unified Medical Language System) Metathesaurus concepts (e.g., UMLS CUI [Concept Unique Identifier]), preferred term, semantic type, and the sentences and predications themselves (subject, predicate, and object).¹¹ For example, if SemRep sees in an abstract the sentence “It is now clear that there is an urgent need to better understand the roles of anxiety and depression in OSAS.” (PMID: [15533283](#)), the subject would be “depressive disorder”, the predicate would be “coexists with”, and the object would be “sleep apnea, obstructive”.

3. MIMIC-III Clinical Database

The Medical Information Mart for Intensive Care III (MIMIC-III) database contains de-identified data for over 40,000 critical care patients from Beth Israel Deaconess Medical Center, located in Boston, MA. It includes demographics, vital signs, laboratory tests, medications, and more.¹¹ MIMIC-III organizes data into tables in which columns contain identifiers and other constant information (e.g., patient ID) and rows contain instantiations of that information. One particularly important table for our study is DIAGNOSES_ICD, which contains a list of all diagnoses associated with a patient. The diagnoses are hospital assigned, and are coded using the International Classification of Disease (ICD), specifically ICD-9-CM.¹²

Data Analysis

The five major steps involved in analyzing MEDLINE/PubMed, SemMedDB, and MIMIC-III were: (1) data selection, (2) data filtering, (3) terminology mapping, (4) generation of basic statistics, and (5) data interpretation and evaluation. The Julia programming language, SQL queries, and Web services were used for the three analyses.

In the MEDLINE/PubMed analysis, data selection involved searching and retrieving MEDLINE records associated with OSA from PubMed. We used the query (performed on July 10, 2016): “sleep apnea, obstructive”[majr], which resulted in 12,550 articles in MEDLINE format. Data filtering involved filtering MeSH descriptors by UMLS semantic type. To do this, we used the NCBI E-utilities EFetch and ESearch in a Julia script and parsed out the MeSH descriptors that were assigned the UMLS semantic type T048 for “Mental or Behavioral Dysfunction”. Terminology mapping required conversion of MeSH descriptors to UMLS CUIs. This was accomplished by direct mapping through the UMLS Terminology Services (UTS) Metathesaurus Browser.

For the SemMedDB analysis, data selection involved using SemMedDB (latest version stored in a local MySQL database) to extract relevant information for our analysis. Data filtering involved finding relationships between OSA and a psychiatric condition (i.e., indicated by a UMLS semantic type of “mobd” for “Mental or Behavioral Dysfunction”). We used the table PREDICATION_AGGREGATE because it contained all relevant information for PubMed citations and semantic predications. *s_name* and *o_name* refer to the subject and object of a semantic predication, and we required one of them to be OSA (for the subject or object). *s_type* and *o_type* refer to the UMLS semantic type of the subject and object, respectively, and one of them was required to be a mental or behavioral dysfunction (i.e., “mobd”). Finally, the predicate connecting the object and the subject was limited to ‘COEXISTS_WITH’, indicating that there is an association between the two. Based on these requirements, the MySQL query used was:

```
select * from PREDICATION_AGGREGATE where (s_name = 'sleep apnea, obstructive' OR o_name = 'sleep apnea, obstructive') AND (s_type = 'mobd' OR o_type = 'mobd') AND predicate = 'COEXISTS_WITH'.
```

No terminology mapping was necessary for SemMedDB because results were already represented as UMLS concepts.

In the MIMIC-III analysis, data selection and filtering involved identifying OSA patients using ICD-9-CM code 327.23 and extracting the unique list of ICD-9-CM codes for each of the 1,768 OSA patients. Like the MEDLINE/PubMed analysis, direct mapping of ICD-9-CM codes to UMLS concepts was accomplished through the UTS Metathesaurus Browser. To more easily classify data, we used the multi-level Clinical Classifications Software (CCS) categories, which is a tool for clustering patient diagnoses and procedures into a manageable number of clinically meaningful categories.¹³ We then calculated frequencies and percentages for each multi-level CCS category.

Data Interpretation and Evaluation

For each dataset, basic statistics were calculated for ranking comorbidities based on frequency and prevalence. Comorbidity rank (CR) was determined using a formula similar to term frequency-inverse document frequency (TF-IDF¹⁴) that is often used in information retrieval to reflect importance of a term relative to a document in a collection and adapt in prior studies¹⁵⁻¹⁶:

$$CR(d_x^c) = \frac{d_x^c}{\sum d_x} \times \log \left(\frac{\sum D}{D^c} \right)$$

where d_x^c is the number of articles (MEDLINE/PubMed), predications (SemMedDB), or patients (MIMIC-III) including disease x (i.e., OSA) and comorbidity c , $\sum d_x$ is the number of articles, predications, or patients including disease x , $\sum D$ is the number of articles, predications, or patients including all diseases (i.e., total number of cases), and D^c is the number of articles, predications, or patients including comorbidity c .

We then used Cohen's kappa as a measure of agreement between the MEDLINE/PubMed and MIMIC-III results. Cohen's kappa coefficient is a statistic that measures inter-rater agreement for qualitative items.¹⁷ The value of kappa is defined as:

$$kappa = \frac{p_0 - p_e}{1 - p_e}$$

where p_0 = agreement and p_e = estimation.

Results

A summary of results for each data source is provided in Table 1.

Table 1. Summary of results.

	MEDLINE/PubMed	SemMedDB	MIMIC-III
Total number	12,550 articles	72 predications	1,768 patients
Total number of psychiatric comorbidities	841	72	1,865
Number of unique psychiatric comorbidities	68	28	283
Highest ranking psychiatric comorbidity (UMLS CUI)	Cognitive deficit (C0009241) ^a	Cognitive deficit (C0009241)	Depressive disorder, NEC (C0868892)

^aCognitive deficit was determined by excluding "Sleep Wake Disorders" and "Mental Disorders" because they did not fit the requirements of a singular psychiatric disorder.

Table 2 includes the 30 MEDLINE/PubMed psychiatric comorbidities with the highest CR values, all over 0.001. Table 3 displays all 28 unique psychiatric comorbidities in SemMedDB and their CR values, also over 0.001. Table 4 includes the 30 MIMIC-III psychiatric comorbidities with the highest CR values. The CR values for MIMIC-III were considerably higher than the other data sources, but only the top 30 are shown for consistency.

Since diagnoses in MIMIC-III were represented as ICD-9-CM codes (prior to mapping to UMLS concepts), this allowed for us to organize the mental disorders according to CCS categories. The CCS for ICD-9-CM is a categorization scheme of over 14,000 diagnosis codes that can be collapsed into smaller and more meaningful categories. For instance, alcohol abuse and alcohol withdrawal would both fit under “Alcohol-related disorders”, which may simplify descriptive statistics and make comparative analyses more effective. Table 5 shows the frequencies of each multi-level CCS category.

Table 2. Top 30 psychiatric comorbidities from MEDLINE/PubMed (as of 2016-08-11)

UMLS CUI	UMLS Preferred Term	CR
C4042891	Sleep Wake Disorders	2.5612
C0004936	Mental disorders	0.1429
C0009241	Cognitive deficit	0.0343
C4041080	Neurocognitive Disorders	0.0320
C0011570	Mental Depression	0.0214
C3714756	Intellectual Disability	0.0173
C0525045	Mood Disorders	0.0149
C0011581	Depressive disorder	0.0129
C0033953	Psychosexual Disorders	0.0123
C0877792	Sleep Disorders, Circadian Rhythm	0.0098
C0025261	Memory Disorders	0.0081
C1321905	Attention deficit hyperactivity disorder	0.0079
C0008066	Child Behavior Disorders	0.0062
C0003469	Anxiety Disorders	0.0054
C0038436	Stress Disorders, Post-Traumatic	0.0043
C0497327	Dementia	0.0038
C0038443	Stress, Psychological	0.0036
C1269683	Major Depressive Disorder	0.0031
C0236969	Substance-Related Disorders	0.0028
C0023186	Learning Disorders	0.0027
C0005586	Bipolar Disorder	0.0025
C0036341	Schizophrenia	0.0023
C0011206	Delirium	0.0023
C0003910	Articulation Disorders	0.0021
C0752299	Sleep-Wake Transition Disorders	0.0021
C1270972	Mild cognitive disorder	0.0018
C0008074	Child Development Disorders, Pervasive	0.0014
C4042784	Feeding and Eating Disorders	0.0014
C0001973	Alcoholic Intoxication, Chronic	0.0014
C0042693	Violence	0.0014

Table 3. All 28 psychiatric comorbidities from SemMedDB (latest version as of June 30, 2015)

UMLS CUI	UMLS Preferred Term	CR
C0009241	Cognitive deficit	0.0119
C0011581	Depressive disorder	0.0092
C0011206	Delirium	0.0041
C0497327	Dementia	0.0030
C2063866	Treatment Resistant Depression	0.0027
C1112442	Female sexual dysfunction	0.0026
C0233794	Memory impairment	0.0026
C0033975	Psychotic Disorders	0.0025
C0033953	Psychosexual Disorders	0.0023
C1306597	Psychiatric problem	0.0020
C0086168	Dissociation	0.0019
C0038436	Stress Disorders, Post-Traumatic	0.0018
C0005586	Bipolar Disorder	0.0017
C0004936	Mental disorders	0.0016
C0036341	Schizophrenia	0.0014
C0856975	Autistic behaviour	0.0013
C0281902	Maladjustment	0.0013
C0849888	Disturbance, psychological	0.0013
C0242151	Violent	0.0013
C0086133	Depressive Syndrome	0.0013
C0085633	Mood swings	0.0012
C0233514	Abnormal behavior	0.0012
C0013473	Eating Disorders	0.0010
C0338831	Manic	0.0010
C0003469	Anxiety Disorders	0.0010
C1321905	Attention deficit hyperactivity disorder	0.0010
C0525045	Mood Disorders	0.0010
C0439857	Dependence	0.0010

Table 4. Top 30 psychiatric comorbidities in MIMIC-III (Version 1.3)

UMLS CUI	UMLS Preferred Term	CR
C0868892	Depressive disorder, NEC	0.2059
C0040335	Encounter due to tobacco use	0.1797
C0040336	Tobacco use disorder	0.1326
C1456303	Transient mental disorders due to conditions classified elsewhere	0.1033
C0700613	Anxiety state	0.0996
C0013415	Dysthymic disorder	0.0923
C0005586	Bipolar Disorder	0.0544
C0038436	Posttraumatic stress disorder	0.0392

C0154326	Drug-induced delirium	0.0334
C0085762	Alcohol abuse	0.0286
C0154338	Other persistent mental disorders due to conditions classified elsewhere	0.0253
C0154475	Other and unspecified alcohol dependence, continuous	0.0252
C0154474	Other and unspecified alcohol dependence, unspecified	0.0249
C0236663	Alcohol withdrawal syndrome	0.0239
C0878691	Dementia in conditions classified elsewhere without behavioral disturbance	0.0229
C0023891	Liver Cirrhosis, Alcoholic	0.0223
C0236773	Depressed bipolar I disorder	0.0220
C3161331	Unspecified intellectual disabilities	0.0217
C0154477	Other and unspecified alcohol dependence, in remission	0.0213
C0375162	Schizo-affective type schizophrenia, unspecified state	0.0194
C0154471	Acute alcoholic intoxication, continuous drinking behavior	0.0188
C0236794	Panic disorder without agoraphobia	0.0180
C0009171	Cocaine abuse	0.0155
C0302371	Other and unspecified special symptoms or syndromes, NEC	0.0154
C0154478	Opioid type dependence, continuous use	0.0154
C0024517	Single major depressive episode	0.0154
C0001539	Adjustment disorder with depressed mood	0.0152
C0036341	Schizophrenia	0.0150
C1456307	Bipolar I disorder, most recent episode (or current) unspecified	0.0149
C0154516	Alcohol abuse, in remission	0.0148

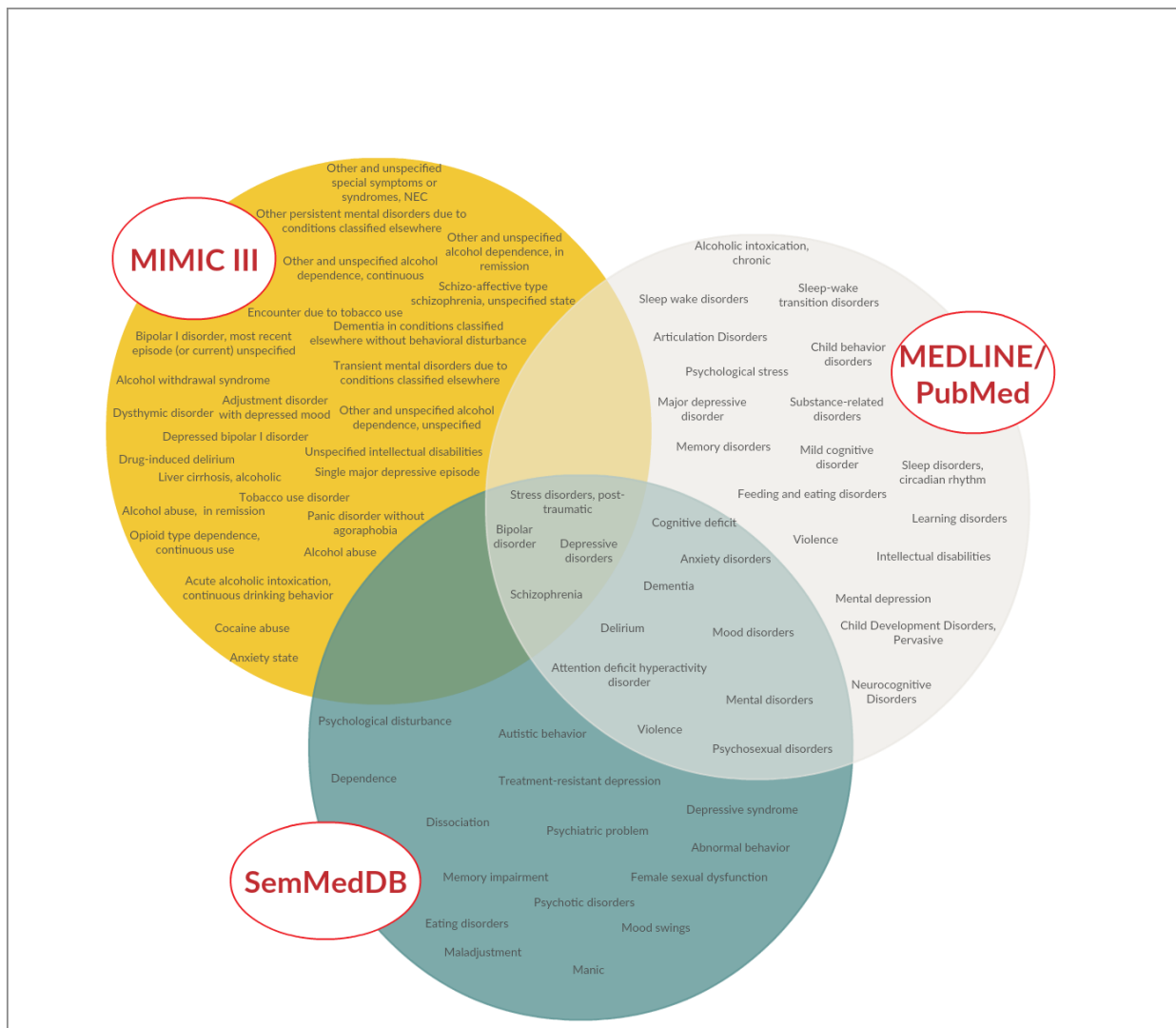
Table 5. Psychiatric comorbidities grouped by multi-level CCS categories for MIMIC III.

CCS Category	Frequency	Relative Frequency (%)
Alcohol-related disorders	6596	21.26
Screening and history of mental health and substance abuse codes	6343	20.35
Mood disorders (including depression)	5915	18.97
Delirium, dementia, and amnesic and other cognitive disorders	4465	14.32
Substance-related disorders	2922	9.37
Anxiety disorders	2168	6.95
Suicide and intentional self-inflicted injury	877	2.81
Schizophrenia and other psychotic disorders	849	2.72
Developmental disorders	494	1.58
Attention-deficit, conduct, and disruptive behavior disorders	198	0.64
Personality disorders	165	0.53
Adjustment disorders	145	0.47
Disorders usually diagnosed in infancy, childhood, or adolescence	31	0.10
Impulse control disorders, NEC	7	0.02

The results of all three analyses were tabulated to determine and compare whether a source found evidence to support an association between OSA and a particular mental disorder. Figure 1 visually depicts the results for MEDLINE/PubMed, SemMedDB, and MIMIC-III (only depicts top ranking comorbidities from Tables 2, 3, and 4). In the full dataset, there were 130 comorbidities in MIMIC-III and 51 comorbidities in MEDLINE/PubMed that had no overlap with other sources. Conduct disorder was the only morbidity found explicitly between MIMIC-III (ranked 109 out of 136 with CR of 0.0021) and MEDLINE/PubMed (ranked 54 out of 68 with CR of 0.0003), and there were no morbidities explicitly shared between SemMedDB and MIMIC-III. However, there was a significant amount of overlap between SemMedDB and MEDLINE/PubMed, which supports SemMedDB being used as a data mining system for MEDLINE biomedical literature. Some of these morbidities include cognitive deficits, depressive disorders, anxiety disorders, and psychosexual disorders. Finally, there was some overlap between the three data sources. Bipolar disorder, depressive disorder, and PTSD were found in all three sources, which verifies previous research. In addition, schizophrenia and other psychotic disorders appeared in all three sources, albeit at lesser significance.

When we compared the level of agreement between MIMIC-III and MEDLINE/PubMed, we found that Cohen's kappa equaled 0.0375, indicating that 3.7% of the time the two sources are in agreement. This is a slight value, but was likely resultant of MIMIC-III having much more specificity through the ICD-9-CM codes whereas the MeSH descriptors in MEDLINE/PubMed were more general.

Figure 1. Venn Diagram depicting the overlap between the three data sources.



Based on these findings, OSA appears to be associated with cognitive deficits, post-traumatic stress disorders, depressive disorders, bipolar disorder, anxiety disorders, mood disorders, alcohol and drug abuse, delirium, dementia, ADHD, schizophrenia, and violence or aggression. There is weaker evidence that OSA is correlated with abnormal behavior, autism, dissociation, psychological disturbance, dysthymic disorder, eating disorders, female sexual dysfunction, maladjustment, mania, psychosexual disorders, suicidal ideation, and treatment-resistant depression.

Discussion

Existing research is aware of the relationship between OSA and mental disorders, particularly depression, anxiety (i.e., sympathetic overactivity), and impaired attention and executive function¹⁸⁻²⁰. OSA leads to frequent sleep disruption, preventing entrance into slow wave (deep) sleep and thus inhibiting alertness and cognition.¹ In addition, due to the repeated shock of being woken abruptly every few minutes, OSA patients have more variable heart rate, higher blood pressure, and increased risk of stroke and heart attack (2-3 fold increase if untreated).¹ These episodes can be frightening and disheartening for patients, which can affect anxiety and depression levels during the daytime. As shown in Figure 1, our results were consistent with existing research, and highlight the prevalence of depression, anxiety, and attention deficit in OSA.

There are admittedly some limitations to consider in all of our sources (MEDLINE/PubMed, SemMedDB, and MIMIC-III). When looking at MeSH descriptors in MEDLINE/PubMed, albeit the minority, sometimes a MeSH descriptor is included in a MEDLINE article to point out its irrelevance (e.g., cognitive deficits that do *not* occur in children may still have “child” as a MeSH descriptor). In addition to analyzing MeSH descriptors²¹, we might consider mapping free-text phrases in titles and abstracts to UMLS concepts (e.g., using the MetaMapped MEDLINE Baseline database) in the future to complement mining the MeSH metadata field, and comparing those results in future research. SemMedDB extracts predication triples (subject-predicate-object) from MEDLINE/PubMed titles and abstracts, ensuring a relationship between the subject (e.g., OSA) and the object (e.g., psychiatric comorbidity). In order to confirm associations, additional predicates could be explored in addition to ‘COEXISTS_WITH’ (e.g., ‘CAUSES’, ‘ASSOCIATED_WITH’, ‘AFFECTS’, ‘INTERACTS_WITH’, or ‘OCCURS_IN’). To eliminate potential confounding predications, negative predicates can also be examined (e.g., ‘NEG_ASSOCIATED_WITH’ and ‘NEG_COEXISTS_WITH’). MIMIC-III includes EHR data, including ICD-9-CM diagnosis codes, for over 40,000 critical care patients from Beth Israel Deaconess Medical Center, and thus provides a clinical perspective through patient-level data. However, utilizing three different types of data sources aims to address the potential biases in each and demonstrate their complementary nature for creating a more well-balanced analysis.

A main challenge was comparison of the data sources due to different levels of granularity. Although comparison between SemMedDB and MEDLINE/PubMed was feasible due to similar granularity levels, comparing either source with MIMIC-III was difficult without grouping due to the fine granularity of the ICD-9-CM codes. For instance, the MIMIC-III data included 20 alcohol-related disorders, some of which were extremely similar, to describe alcohol abuse. These UMLS CUIs include C0154473 for “Acute alcoholic intoxication in remission, in alcoholism” and C1812624 for “Acute alcoholic intoxication in alcoholism, continuous drinking behavior”. Because of this granularity, many-to-one mapping could be performed to collapse the data into smaller and more meaningful categories. One option is to use existing categorizations such as the single- and multi-level CCS categories described above (Table 4) or hierarchical relationships defined in the UMLS Metathesaurus. Another advantage to grouping these data would likely be an increase in CR values, since it would increase d_x^c for many morbidities. Next steps include to applying these groupings to each of the datasets and conducting a comparison with more similar granularity levels. This would potentially also enable us to better assess kappa, or the agreement among MEDLINE/PubMed, SemMedDB, and MIMIC-III. We anticipate that after grouping ICD-9-CM codes and MeSH descriptors using different categorization schemes (e.g., single- and multi-level CCS categories), more parallel comparisons could be conducted leading to higher agreement between the sources.

Importantly, what our preliminary analysis found is that prior research often overlooks the high prevalence of cognitive deficits in the OSA population. Indeed, cognitive deficit (UMLS CUI C0009241) occurred the most frequently alongside OSA in MEDLINE/PubMed and SemMedDB. Furthermore, as Tables 2 and 3 show, there were numerous other psychiatric comorbidities in MEDLINE/PubMed and SemMedDB with high comorbidity ranks (i.e., CRs around 0.02 or above) that could be classified as cognitive disorders, such as neurocognitive disorders (C4041080), memory disorders (C0025261), dementia (C0497327), learning disorders (C0023186), delirium (C0011206), and mild cognitive disorder (C1270972). Table 5 shows that in addition, delirium, dementia,

and amnesic and other cognitive disorders made up 14.32% of all comorbidities in MIMIC-III. After grouping (e.g., using CCS), we anticipate finding high overlap of various cognitive disorders/deficits among all three data sources.

Cognition disorders and cognitive deficits likely arise from the severe sleep fragmentation and reduced total sleep time in the OSA population, as sleep has been proven to be vital for attention, information consolidation, memory encoding and retrieval, and executive functioning.¹ With apnea drastically impairing sleep health, it is no surprise that these functions would be subsequently incapacitated. Patients with OSA frequently have neurocognitive deficits, particularly in regards to selective attention and concentration, sustained attention, short-term and working memory, and executive function.²² Although some of these neurocognitive deficits might be attenuated with proper treatment, executive dysfunction is persistent and may continue to impact quality of life.²² Consequently, more research is needed to determine how to prevent OSA-related cognitive decline in at-risk of afflicted populations.

Next steps include formally evaluating the results for specific mental disorders (e.g., depressive disorders) with clinical experts and other medical knowledge sources, and continuing to create and validate ranked metrics to better assess the associations. Other future work could involve generating association rules and sequential patterns to study non-temporal and temporal relationships between multiple psychiatric conditions for OSA, as well as looking at differences in psychiatric comorbidities based on age, gender, and race/ethnicity. Furthermore, the approach used did not account for confounding information (e.g., the high prevalence of depression worldwide and whether or not it is explicitly associated with OSA), which could lead to high false positive discovery rates. Consequently, next steps will aim to reduce confounders by determining psychiatric comorbidity rates in various demographic groups and potentially comparing OSA patients with the general population.

Conclusion

The association between OSA and mental disorders appears strong, and this study provides additional evidence that mental and behavioral health should be assessed and treated alongside OSA to improve patient quality of life. One main finding of our research is the significant relationship between OSA and cognitive deficits, as evidence by analysis in all three data sources. Because OSA impairs sleep hygiene, we postulate that the cognitive benefits of sleep are diminished and persons with OSA, and deficits may develop if the sleep debt is chronic. Indeed, OSA is simultaneously a determinant of health and an outcome of impaired health, and having enhanced understanding of OSA comorbidities using biomedical and clinical data sets could be a valuable tool in measuring and monitoring health.

Acknowledgments

The authors thank Ashley Lee for contributing to retrieval of MIMIC-III data and Isabel Restrepo and Alice Chu for discussions related to the MEDLINE/PubMed analysis. This work was supported in part by National Library of Medicine grant R01LM011364.

References

1. Lockley SW, Foster RG. Sleep: a very short introduction. Oxford: Oxford University Press; 2012.
2. Huang Q, Qin Z, Zhang S, Chow C. Clinical patterns of obstructive sleep apnea and its comorbid conditions: a data mining approach. *J Clin Sleep Med*. 2008;4(6):543-550.
3. Sharafkhaneh A, Giray N, Richardson P, Young T, Hirshkowitz M. Association of psychiatric disorders and sleep apnea in a large cohort. *SLEEP*. 2005;28(11): 1405-1411.
4. Gupta MA, Simpson FC. Obstructive sleep apnea and psychiatric disorders: a systematic review. *J Clin Sleep Med* 2015;11(2):165–175.
5. U.S. National Library of Medicine Fact Sheets. Available from: <http://www.nlm.nih.gov/pubs/factsheets/>
6. Lindberg DA, Humphreys BL. Rising expectations: access to biomedical information. *Yearb Med Inform*. 2008;165-72.
7. Zhang Y, Sarkar IN, Chen ES. PubMedMiner: Mining and visualizing MeSH-based associations in PubMed. *AMIA Annu Symp Proc*. 2014:1990–1999.
8. Kilicoglu H, Shin D, Fiszman M, Rosemblat G, Rindfleisch TC. SemMedDB: a PubMed-scale repository of biomedical semantic predications. *Bioinformatics*. 2012 Dec 1;28(23):3158-60.
9. SemMedDB details. Available from: <https://skr3.nlm.nih.gov/SemMedDB/dbinfo.html>
10. Rindfleisch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *J Biomed Inform*. 2003;36(6):462-477.
11. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016 May 24;3:160035.

12. Overview of the MIMIC-III data. Available from: <https://mimic.physionet.org/gettingstarted/overview/>
13. Elixhauser A, Steiner CA, Whittington C, et al. Clinical classifications for health policy research: Hospital inpatient statistics, 1995. Healthcare Cost and Utilization Project, HCUP 3 Research Note. Rockville, MD: Agency for Health Care Policy and Research; 1998. AHCPR Pub. No. 98-0049.
14. Salton G, Fox EA, Wu H. Extended Boolean information retrieval. *Communications of the ACM*. 1983;26(11):1022-1036.
15. Sarkar IN, Chen ES. Determining compound comorbidities for heart failure from hospital discharge data. *AMIA Annu Symp Proc*. 2012;2012:809-18.
16. Chen ES, Melton GB, Wasserman RC, Rosenau PT, Howard DB, Sarkar IN. Mining and Visualizing Family History Associations in the Electronic Health Record: A Case Study for Pediatric Asthma. *AMIA Annu Symp Proc*. 2015 Nov 5;2015:396-405.
17. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia medica*. 2012;22(3):276.
18. Caples SM, Gami AS, Somers VK. Obstructive sleep apnea. *Ann Intern Med*. 2005; 142:187-197.
19. Ip MSM, Lam B, Ng MMT, Lam WK, Tsang KWT, Lam KSL. Obstructive sleep apnea is independently associated with insulin resistance. *Am J Respir Crit Care Med*. 2002; 165:670–676.
20. Epstein LJ, Kristo D, Strollo PJ, Friedman N, Malhotra A, Patil SP, Ramar K, Rogers R, Schwab RJ, Weaver EM, Weinstein MD. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med*. 2009; 5(3):263-276.
21. Srinivasan P. Text mining: generating hypotheses from MEDLINE. *J Am Soc Inf Sci*. 2004;55(5):396-413.
22. Alchanatis M, Zias N, Deligiorgis N, Amfilochious A, Dionellis G, Orphanidou D. Sleep apnea-related cognitive deficits and intelligence: an implication of cognitive reserve theory. *J Sleep Research*. 2005;14:69-75.