

SCIENTIFIC INVESTIGATIONS

Multisite validation of a simple electronic health record algorithm for identifying diagnosed obstructive sleep apnea

Brendan T. Keenan, MS^{1,*}; H. Lester Kirchner, PhD^{2,*}; Olivia J. Veatch, PhD^{1,3}; Kenneth M. Borthwick, MS²; Vicki A. Davenport, HSD²; John C. Feemster, BA⁴; Maged Gendy, BS⁵; Thomas R. Gossard, BA⁴; Frances M. Pack, RN¹; Laura Sirikulvadhana, MPH⁶; Luke N. Teigen, BS⁴; Paul C. Timm, MNM⁴; Beth A. Malow, MD, MS³; Timothy I. Morgenthaler, MD⁴; Phyllis C. Zee, MD, PhD⁵; Allan I. Pack, MBChB, PhD¹; Janet D. Robishaw, PhD⁷;‡; Stephen F. Derose, MD, MS⁶;†

¹Division of Sleep Medicine/Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; ²Biomedical and Translational Informatics, Geisinger, Danville, Pennsylvania; ³Sleep Disorders Division, Vanderbilt University Medical Center, Nashville, Tennessee; ⁴Center for Sleep Medicine, Mayo Clinic, Rochester, Minnesota; ⁵Center for Circadian and Sleep Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; ⁶Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, California; ⁷Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, Florida; *Co-lead authors; †Joint senior authors

Study Objectives: We examined the performance of a simple algorithm to accurately distinguish cases of diagnosed obstructive sleep apnea (OSA) and noncases using the electronic health record (EHR) across six health systems in the United States.

Methods: Retrospective analysis of EHR data was performed. The algorithm defined cases as individuals with ≥ 2 instances of specific International Classification of Diseases (ICD)-9 and/or ICD-10 diagnostic codes (327.20, 327.23, 327.29, 780.51, 780.53, 780.57, G4730, G4733 and G4739) related to sleep apnea on separate dates in their EHR. Noncases were defined by the absence of these codes. Using chart reviews on 120 cases and 100 noncases at each site (n = 1,320 total), positive predictive value (PPV) and negative predictive value (NPV) were calculated.

Results: The algorithm showed excellent performance across sites, with a PPV (95% confidence interval) of 97.1 (95.6, 98.2) and NPV of 95.5 (93.5, 97.0). Similar performance was seen at each site, with all NPV and PPV estimates ≥ 90% apart from a somewhat lower PPV of 87.5 (80.2, 92.8) at one site. A modified algorithm of ≥ 3 instances improved PPV to 94.9 (88.5, 98.3) at this site, but excluded an additional 18.3% of cases. Thus, performance may be further improved by requiring additional codes, but this reduces the number of determinate cases.

Conclusions: A simple EHR-based case-identification algorithm for diagnosed OSA showed excellent predictive characteristics in a multisite sample from the United States. Future analyses should be performed to understand the effect of undiagnosed disease in EHR-defined noncases. This algorithm has wide-ranging applications for EHR-based OSA research.

Keywords: case-identification algorithm, diagnostic codes, electronic health record, obstructive sleep apnea, phenotype

Citation: Keenan BT, Kirchner HL, Veatch OJ, et al. Multisite validation of a simple electronic health record algorithm for identifying diagnosed obstructive sleep apnea. *J Clin Sleep Med*. 2020;16(2):175–183.

BRIEF SUMMARY

Current Knowledge/Study Rationale: A simple case-identification algorithm for obstructive sleep apnea using electronic health record data allows researchers to leverage the wealth of healthcare data linked to biorepositories. There are limited data on performance of this type of algorithm. Using clinical chart reviews at six health centers in the United States, we validated an algorithm based on diagnostic codes.

Study Impact: Results demonstrate that an algorithm defining cases as individuals with at least two instances (on separate dates) of diagnostic codes for sleep apnea in their medical record and noncases as individuals with no codes results in high positive and negative predictive values. This algorithm has wide-ranging applications, including monitoring quality and utilization metrics, clinical epidemiology, and genetic association studies.

INTRODUCTION

Obstructive sleep apnea (OSA) is characterized by repetitive episodes of full or partial cessation of breathing and associated decreases in oxygen saturation during sleep.¹ OSA is a common and treatable condition, yet often remains undiagnosed or untreated, and thus carries a large public health burden.² OSA is associated with cardiovascular and cerebrovascular diseases,^{3,4} cancer incidence and mortality,⁵ neurocognitive deficits and sleepiness,⁶ automobile and work accidents,^{7,8} and lower quality of life.⁹ Therefore, establishing a well-performing algorithm to

identify cases of OSA using readily available clinical data has clear public health relevance.

The growth of electronic health records (EHRs) in recent years provides a unique opportunity for large-scale case identification through the application of automated algorithms and allows large-scale clinical research and quality improvement projects using EHR data. To identify patients that are diagnosed with the disease of interest, counts of disease-specific International Classification of Diseases (ICD) diagnostic codes in the EHR are routinely leveraged. Although some challenges have been noted,¹⁰ these codes have the advantage

of being universally and readily available, standardized across health systems, efficient and straightforward to query on large patient populations, and are indicative of specific clinical judgments. Because diagnostic codes at times may be applied prematurely to justify diagnostic testing (ie, presumptive diagnostic coding), methods to improve case identification include requiring multiple instances of codes or combining codes with key clinical information (eg, body mass index [BMI], age, sex) and/or codes for comorbidities related to the primary disease of interest (eg, hypertension, diabetes, cardiovascular diseases). Ultimately, case identification algorithms can be used for a variety of clinical and research purposes, including quality and utilization studies, patient outreach, and clinical study recruitment and outcomes. However, to ensure that case identification algorithms are accurately capturing the disease of interest, validation studies evaluating performance characteristics against gold-standard clinical chart reviews are required.

The current study leverages EHR data from six sites throughout the United States that participate in the Sleep Apnea Genetics Study to develop and validate a simple, diagnostic code-based algorithm for identifying cases of diagnosed OSA. Establishing a well-performing and efficient case identification algorithm has important implications for OSA-related research in large biorepositories, with potential applications ranging from efficient patient outreach to performing large-scale genotype-phenotype association analyses leveraging linked genetic information.

METHODS

Overview and study populations

We performed a retrospective analysis of EHR data from each of the six institutions participating in the Sleep Apnea Genetics Study, namely Geisinger, Kaiser Permanente Southern California, Mayo Clinic, Northwestern University, University of Pennsylvania, and Vanderbilt University Medical Center. Human subjects research approval was obtained from the institutional review board at each site.

EHR data were independently accessed at each site to identify study participants aged 18 to 88 years for clinical chart reviews. An EHR-based algorithm (described in the next paragraphs) was used to categorize this population into cases of diagnosed OSA, noncases, and indeterminate cases. The final study cohort from which a sample was drawn for algorithm validation at each site always included some persons who had contributed samples to a genetic biobank, given the consortium's goal of utilizing the algorithm to perform future genetic association analyses. Most of the sites required all patients to have been consented into genetic repositories, whereas the University of Pennsylvania included patients from the entire health system, a small proportion of whom were consented into the biobank. A sample of cases and noncases was chosen from each site for chart review using a computerized random selection process. The goal of chart review was to determine whether OSA was diagnosed or not based on definitive criteria (described in the next paragraphs), independent from diagnostic codes; we recognize that absence of a diagnosis (eg, noncases) refers only to current diagnostic status, as OSA is a progressive and often underdiagnosed

disorder.¹¹ The supplemental material contains further details on each site's source population and demographic characteristics of the validation sample.

EHR-based diagnostic case identification algorithm

An EHR-based diagnostic case identification algorithm was developed after considering the feasibility and generalizability of various electronic data for identifying diagnosed cases of sleep apnea. We considered sleep study test results with an associated OSA diagnosis or result (eg, elevated apnea-hypopnea index); however, there was clearly a great deal of variation in test result retrieval processes and formatting, which necessitated a chart review approach in most sites. We also considered positive airway pressure (PAP) devices and equipment (eg, masks) with an associated OSA diagnosis; however, several sites did not have sufficient data on durable medical equipment orders or dispensations to make this a viable approach. Thus, we chose to test criteria based solely on instances of ICD-9 and ICD-10 diagnostic codes, which are essentially required for clinical care and reimbursement. Specifically, we hypothesized that cases could be defined as any individual with two or more instances (on separate dates) of any of the following diagnostic codes (description): 327.20 (Organic sleep apnea, unspecified), 327.23 (Obstructive sleep apnea [adult, pediatric]), 327.29 (Other organic sleep apnea), 780.51 (Insomnia with sleep apnea), 780.53 (Hypersomnia with sleep apnea), 780.57 (Sleep apnea [NOS]), G4730 (Sleep apnea, unspecified), G4733 (Obstructive sleep apnea [adult, pediatric]) and G4739 (Other sleep apnea). Although some of these codes may include central sleep apnea, it was thought that most of them would be associated with predominantly OSA and that it was important not to overly restrict case selection by avoiding codes that do not differentiate between central and obstructive. Conversely, we considered that noncases could be defined as any individuals with zero instances of any of the listed diagnostic codes in their medical record. Individuals with a single instance of any of the diagnostic codes were classified as indeterminate and excluded from any further analyses.

Clinical chart review procedures

To determine the accuracy of the proposed EHR-based identification of OSA, medical records from 120 randomly selected EHR-defined cases and 100 EHR-defined noncases were manually reviewed by trained experts at each participating site. A larger number of EHR-positive patients was chosen to maintain an adequate number of cases for assessment of alternative case definitions. Thus, a total of 1,320 randomly chosen participants ($n = 720$ EHR-positive, $n = 600$ EHR-negative) were reviewed to provide estimates of performance characteristics. This sample is similar or larger than validation studies of other new phenotypes conducted at participating sites through collaboration with the eMERGE Network.¹²

When available, true cases were defined using results from diagnostic polysomnography (PSG) or ambulatory home sleep tests; these data were available for 72% of confirmed OSA cases. Cases were confirmed based on the clinician diagnoses included in the sleep study reports (for example, as based on the International Classification of Sleep Disorders [ICSD] criteria of apnea-hypopnea index [AHI] > 5 events/h, regardless of the

specific criteria used to define hypopneas). When sleep tests were not available, cases were confirmed based on any of the following criteria: (1) presence of a diagnosis of OSA either included in the EHR problem list or from clinician notes citing evidence from a sleep study, (2) evidence of prescriptions for, or titration of, continuous positive airway pressure (CPAP, or auto-CPAP) treatment, and/or (3) evidence of other OSA treatments or procedures (eg, mandibular advancement device). Central sleep apnea was acceptable within our defined cases as long as OSA was also diagnosed. Noncases were primarily confirmed by manually searching and reviewing notes to identify references to: (1) sleep disorders, (2) sleep disturbances, (3) snoring, (4) CPAP use, and/or (5) sleep tests or the results of a sleep test (eg, AHI). When these references were identified in a record, approximately two to three sentences of text surrounding the occurrence of the word were evaluated to determine if the context indicated that the individual may be an undiagnosed case of OSA. For some sites, when PSGs were available for identified noncases, the absence of OSA was based on clinician confirmation of no OSA from sleep study reports (ie, $AHI < 5$ events/h). During these chart reviews, information on the electronic and clinical diagnoses, demographics, and disease severity (when sleep studies were available) were collected in a standardized REDCap database.¹³

To ensure accuracy in case or noncase confirmation from clinical chart review methods, 10 randomly chosen EHR-defined cases and 10 randomly chosen EHR-defined noncases were reviewed by a second independent reviewer at each site using the same approach. Reliability was assessed by calculating percentage agreement and kappa coefficients based on clinical diagnoses determined by the two reviewers.

Statistical analysis

Continuous variables are summarized using means and standard deviations and categorical variables using frequencies and percentages. Variables are compared between cases and noncases using *t* tests for continuous variables and chi-square or Fisher exact tests for categorical variables.

To determine the performance of the EHR-based identification algorithm against clinical diagnosis from medical chart reviews, we calculated appropriate measures of positive predictive value (PPV; the probability that an individual has diagnosed OSA given they are identified as a case using the EHR-based algorithm) and negative predictive value (NPV; the probability that an individual does not have diagnosed OSA given they are identified as a noncase using the EHR-based algorithm). To understand the variability of each point estimate, we also calculated the 95% exact binomial confidence intervals (95% CIs) for each proportion. Analyses were performed with all samples combined and stratified by site. Validity of the algorithm was defined *a priori* as values of at least 90% for both NPV and PPV. If performance for the primary case definition fell below this *a priori* criteria, we examined the performance of alternative case definitions (eg, at least three instances of diagnosis codes) to understand the trade-off between case identification criteria performance characteristics and the number of cases recategorized as indeterminate. Given the fixed sampling proportions in our validation study (100 EHR-defined noncases,

120 EHR-defined cases per site), measures of sensitivity and specificity were calculated indirectly as functions of PPV, NPV, and assumed disease prevalences.¹⁴ Where calculated, $P < .05$ was considered statistically significant. Statistical analyses were performed using Stata/SE 14.2 (StataCorp LLC, College Station, Texas, United States).

RESULTS

Reliability of chart review procedures

To measure the reliability of clinical diagnosis from chart reviews, 10 randomly chosen EHR-defined cases and 10 randomly chosen EHR-defined noncases were reviewed by two independent reviewers at each site. A comparison of the clinical diagnoses between the two reviewers resulted in 97.5% agreement and a kappa coefficient of 0.950 pooled across sites (**Table S1** in the supplemental material). Within each site, percent agreement between the two independent reviewers ranged from 95.0% to 100.0% and kappa coefficients ranged from 0.900 to 1.000. Thus, using the aforementioned chart review procedures resulted in reliable clinical diagnoses.

Sample characteristics

For validation of the EHR algorithm, 120 individuals identified as cases and 100 identified as noncases were randomly chosen from each participating site ($n = 1,320$ total, 720 EHR-positive and 600 EHR-negative). Sample characteristics are presented in **Table 1**. On average, the validation sample was middle to older aged (mean \pm standard deviation age of 56.0 ± 14.9 years), were moderately obese (BMI of $33.1 \pm 9.0 \text{ kg/m}^2$), predominantly Caucasian (79.3%) and majority female (57.6%). As expected, there are significant differences in the distribution of demographic characteristics across sites, underscoring the diversity of the participating sites (**Table 1**).

Sample characteristics stratified by EHR-defined status are presented overall (**Table 2**) and separately by site (**Table S2** in the supplemental material). Supporting clinical validity of our proposed EHR algorithm, compared to noncases, patients identified as cases by the algorithm were older (57.6 ± 12.2 versus 54.1 ± 17.3 years; $P < .0001$), more obese (BMI of 36.2 ± 9.2 versus $29.3 \pm 7.0 \text{ kg/m}^2$; $P < .0001$), and more likely to be male (48.1% versus 35.5%; $P < .0001$). Further, when using age, sex, and BMI to calculate the predicted probability of OSA based on the Symptomless Multivariable Apnea Prediction Score¹⁵ (**Figure 1**), clear differences can be seen in the predicted OSA probabilities; approximately 70% of EHR-defined noncases have a predicted OSA probability < 0.2 and only $< 3\%$ have a predicted probability ≥ 0.7 . As shown in **Table S2**, there are some distinct demographic differences (age, BMI, sex, and race/ethnicity) when comparing EHR cases and noncases within each site separately. However, this did not adversely affect performance of the algorithm (detailed in the next paragraphs).

Algorithm performance

Performance characteristics for our proposed EHR-based OSA identification algorithm are presented in **Table 3** and shown in **Figure 2**. Among all included patients, the algorithm showed

Table 1—Demographic characteristics of the validation sample, overall and by site.

Measure	All Sites (n = 1,320)	Geisinger (n = 220)	KPSC (n = 220)	Mayo (n = 220)	NU (n = 220)	Penn (n = 220)	Vanderbilt (n = 220)	P
Male, %	42.4	41.4	45.9	58.2	28.6%	32.9	47.3	< .0001
Age, years	56.0 ± 14.9	53.9 ± 14.9	55.4 ± 15.0	63.7 ± 14.3	57.1 ± 11.9	47.4 ± 15.2	58.7 ± 12.7	< .0001
BMI, kg/m ²	33.1 ± 9.0	35.4 ± 9.3	31.7 ± 7.1	31.6 ± 7.6	32.4 ± 9.2	34.5 ± 10.6	32.7 ± 8.9	< .0001
Race, %								< .0001
Caucasian	79.3	98.2	81.2	97.3	66.3	44.2	85.9	
African American	15.6	1.4	6.1	0.5	23.9	52.1	11.4	
Asian	1.4	0.5	6.1	0.0	0.5	1.4	0.0	
Other	3.7	0.0	6.6	2.3	9.2	2.3	2.7	
Hispanic, %	5.4	0.9	23.3	0.5	6.8	1.4	0.0	< .0001

Data presented as mean ± standard deviation or %. P values derived from analysis of variance or chi-square test comparing values among sites. BMI = body mass index, KPSC = Kaiser Permanente Southern California, NU = Northwestern University, Penn = University of Pennsylvania.

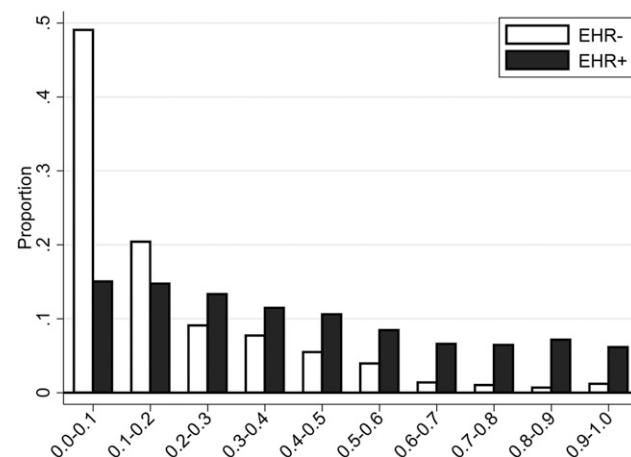
Table 2—Demographic characteristics of electronic health record-based cases and noncases in validation sample.

Measure	Noncases (n = 600)	Cases (n = 720)	P
Male, %	35.5	48.1	< .0001
Age, years	54.1 ± 17.3	57.6 ± 12.3	< .0001
BMI, kg/m ²	29.3 ± 7.0	36.2 ± 9.2	< .0001
Race, %			< .0001
Caucasian	79.9	78.8	
African American	11.3	18.9	
Asian	2.2	0.8	
Other	6.6	1.4	
Hispanic, %	7.5	3.7	.002

Data presented as mean ± standard deviation or %. P values derived from t test and chi-square or Fisher exact tests comparing noncases and cases. BMI = body mass index.

high performance with respect to accurately identifying diagnosed OSA cases and noncases, meeting our *a priori* specified requirement of all values of PPV and NPV of at least 90%. In fact, all estimates exceeded 95% in the full sample, including a PPV (95% CI) of 97.1 (95.6, 98.2) and NPV of 95.5 (93.5, 97.0). Thus, in our multisite sample, patients screening positive using our simple algorithm have a high probability of having diagnosed OSA and patients screening negative have a high probability of not having an OSA diagnosis, based on clinical chart reviews.

We also examined within-site performance characteristics (**Table 3** and **Figure 2**). Nearly all estimates achieved the *a priori* threshold of at least 90%, with the exception of a PPV of 87.5 (80.2, 92.8) at Geisinger. To determine whether alternative definitions are required, and to understand the relative trade-off between performance and proportion of indeterminate patients, we examined changes in PPV and reduction in the number of cases using more stringent case definitions of three or more or four or more instances of diagnostic codes (**Table 4**); NPV remains constant when changing only the case definition. At Geisinger, requiring three instances of diagnostic codes increased

Figure 1—Predicted OSA probabilities in EHR-defined cases and noncases.

The predicted probability of OSA calculated using the Symptomless Multivariable Apnea Prediction Score, a function of age, body mass index, and sex,¹⁵ is presented among EHR-defined cases (EHR+) and noncases (EHR-). Results show the clinical validity of the algorithm, with a clear skew of noncases toward the lower predicted OSA probabilities. Approximately 70% of noncases have a predicted OSA probability below 20%, suggesting a low likelihood of undiagnosed OSA among a majority of the EHR-defined noncases, and less than 5% have a predicted OSA probability above 60%. EHR = electronic health record, OSA = obstructive sleep apnea.

the PPV from 87.5 (80.2, 92.8) to 94.9 (88.5, 98.3), but reclassified 22 patients (18.3%) with exactly two diagnostic code instances as indeterminate. At other participating sites (**Table 4**), comparable or higher PPV were found when requiring ≥ 3 instances of diagnostic codes, with reductions in the number of determinate cases ranging from 7.5% (at Vanderbilt University) to 46.7% (at University of Pennsylvania). Similar results were seen when requiring four or more instances of diagnostic codes (**Table 4**) when compared to the original definition, ranging from 15.8% (at Vanderbilt University) to 79.2% (at University of Pennsylvania). Thus, when considering alternative definitions, the relative improvement in PPV should be weighed

Table 3—Performance characteristics of the electronic health record algorithm overall and within each site.

Site	PPV	NPV
All participants	97.1 (95.6, 98.2)	95.5 (93.5, 97.0)
Geisinger	87.5 (80.2, 92.8)	95.0 (88.7, 98.4)
KPSC	100.0 (97.0, 100.0)	99.0 (94.6, 100.0)
Mayo Clinic	100.0 (97.0, 100.0)	98.0 (93.0, 99.8)
NU	98.3 (94.1, 99.8)	90.0 (82.4, 95.1)
Penn	99.2 (95.4, 100.0)	97.0 (91.5, 99.4)
Vanderbilt	97.5 (92.9, 99.5)	94.0 (87.4, 97.8)

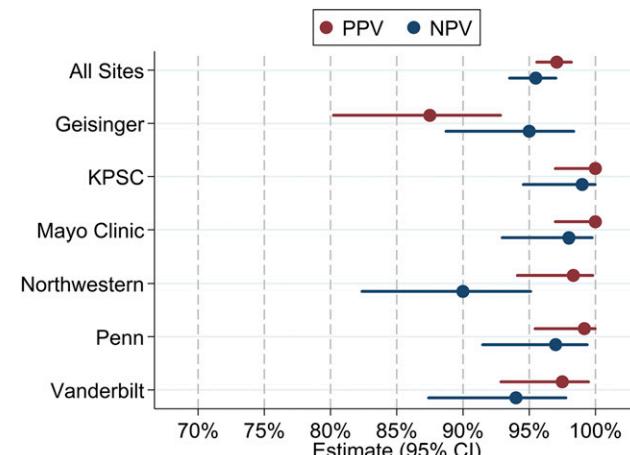
Data presented as estimate (95% exact confidence interval). KPSC = Kaiser Permanente Southern California, NPV = negative predictive value, NU = Northwestern University, Penn = University of Pennsylvania, PPV = positive predictive value.

against the clear reduction in defined cases. Ultimately, a definition of at least two instances appears optimal at nearly all sites.

Sensitivity and specificity estimates derived from observed PPV and NPV and assumed prevalence values are presented in **Figure 3** for OSA prevalences ranging from 15% to 50%; data from the Wisconsin Sleep Cohort suggest an OSA prevalence (defined as AHI ≥ 5 events/h) of 17.4% among middle-aged women and 33.9% among middle-aged men.¹⁶ As assumed OSA prevalence increases, estimated sensitivity increases and specificity decreases. Among the full sample, our algorithm achieves a sensitivity of 77.7% and specificity of 99.5% at the assumed prevalence of 17.4% in women and a sensitivity of 90.9% and specificity of 98.6% at the assumed prevalence of 33.9% in men.

DISCUSSION

OSA is a common condition, and the full consequences of OSA and its treatments (eg, positive airway pressure therapy) are still under investigation. Many research and clinical applications leveraging existing EHR databases can be enhanced by validated case identification criteria. Based on our evaluation of the performance of an algorithm using instances of diagnostic codes for sleep apnea in the EHR, individuals and populations with diagnosed OSA can be accurately identified at six clinical sites in the United States. Across all sites, the simple algorithm of ICD-9/ICD-10 diagnostic codes for sleep apnea on at least two separate dates in an individual's health record produced positive and negative predictive values over 95% for diagnosed OSA. Similarly high performance of this algorithm was found within each site; one site (Geisinger) with a slightly lower PPV achieved performance similar to other sites by requiring one additional instance of a diagnosis code (eg, on at least three different dates). However, this alternative definition resulted in a higher proportion of indeterminate diagnoses, highlighting the importance of considering the appropriate balance between accuracy and sample size. This case identification algorithm has potential wide-ranging applications.

Figure 2—Performance characteristics of primary EHR-based OSA identification algorithm.

Performance characteristics (PPV and NPV) against gold-standard clinical chart review for an EHR-based OSA case definition of two or more ICD-9/ICD-10 diagnostic codes related to OSA is shown overall and at each participating site. CI = confidence interval, EHR = electronic health record, ICD = International Classification of Diseases, KPSC = Kaiser Permanente Southern California, NPV = negative predictive value, OSA = obstructive sleep apnea, Penn = University of Pennsylvania, PPV = positive predictive value.

The proposed algorithm has utility for both clinical and research purposes. Clinical applications include monitoring quality and utilization metrics, and risk assessment with triage of patients for targeted services, including outreach from population health care programs. For example, clinical managers can utilize the algorithm to track patients in their health systems who have OSA and assess their disease outcomes (eg, cardiovascular or metabolic comorbidities) before and after diagnosis. This type of longitudinal tracking is particularly relevant at institutions with stable patient populations, such as Geisinger, where there is less than 1% migration of patients out of the system per year (<https://www.geisinger.edu/research/research-and-innovation/resources>). This application will be further strengthened by obtaining data on ongoing treatment (eg, PAP). Metrics on PAP adherence and efficacy (eg, h/night, residual AHI, mask leak) can now be obtained remotely from modern machines, allowing application of data analytical strategies to evaluate temporal changes in adherence.^{17–20}

Clinicians can also use this algorithm to identify those “noncases” most likely to have undiagnosed OSA. In our sample, most noncases had low probability of OSA using the Symptomless Multivariable Apnea Prediction Score, based on age, sex, and BMI.¹⁵ However, a small proportion of noncases (2.9%) had a high probability of OSA (ie, ≥ 0.7). These noncases at high risk could be identified, contacted, and further studied to determine whether they have undiagnosed OSA. This identification of noncases at high risk of OSA can also be combined with other data in the EHR to inform future studies targeting OSA as a modifiable risk factor for prevention of other adverse outcomes. For example, OSA may accelerate progression of neurodegeneration.^{21,22} Thus, early

Table 4—PPV and remaining cases for alternative electronic medical record case definitions.

Site	PPV	Remaining Cases
At least 2 diagnostic code instances		
All participants	97.1 (95.6, 98.2)	720 (100.0)
Geisinger	87.5 (80.2, 92.8)	120 (100.0)
KPSC	100.0 (97.0, 100.0)	120 (100.0)
Mayo Clinic	100.0 (97.0, 100.0)	120 (100.0)
NU	98.3 (94.1, 99.8)	120 (100.0)
Penn	99.2 (95.4, 100.0)	120 (100.0)
Vanderbilt	97.5 (92.9, 99.5)	120 (100.0)
At least 3 diagnostic code instances		
All participants	98.4 (97.0, 99.3)	573 (79.6)
Geisinger	94.9 (88.5, 98.3)	98 (81.7)
KPSC	100.0 (96.5, 100.0)	105 (87.5)
Mayo Clinic	100.0 (96.4, 100.0)	102 (85.0)
NU	100.0 (96.1, 100.0)	93 (77.5)
Penn	98.4 (91.6, 100.0)	64 (53.3)
Vanderbilt	97.3 (92.3, 99.4)	111 (92.5)
At least 4 diagnostic code instances		
All participants	99.0 (97.6, 99.7)	488 (67.8)
Geisinger	98.9 (94.0, 100.0)	90 (75.0)
KPSC	100.0 (96.2, 100.0)	95 (79.2)
Mayo Clinic	100.0 (96.2, 100.0)	96 (80.0)
NU	100.0 (95.5, 100.0)	81 (67.5)
Penn	96.0 (79.6, 99.9)	25 (20.8)
Vanderbilt	97.0 (91.6, 99.4)	101 (84.2)

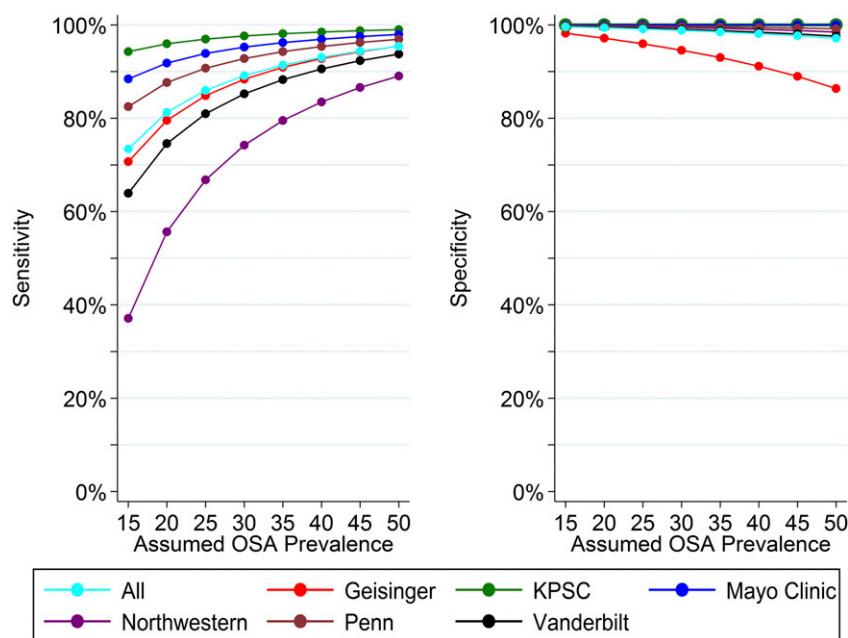
PPV presented as estimate (95% exact confidence interval). Remaining cases presented as n (%). KPSC = Kaiser Permanente Southern California, NU = Northwestern University, Penn = University of Pennsylvania, PPV = positive predictive value.

identification and treatment of OSA in patients at increased risk of neurodegenerative disorders (eg, Alzheimer disease) could provide therapeutic benefits. This needs to be assessed in future studies.

Beyond these immediate clinical applications, the case identification algorithm can be combined with data in the EHR to enhance research projects. When combined with available information on treatment adherence (noted previously), outcomes among OSA cases adherent to PAP compared to those who are not can be analyzed. These analyses can then be combined with EHR data to predict the subsets of patients in whom OSA was diagnosed likely to demonstrate the greatest benefits of treatment. By applying the algorithm to health systems with large biobanks and genotype data, as in the current manuscript, genetic analyses (eg, genome-wide association studies) can be applied without the need to recruit new patients, thereby considerably reducing the cost of such studies. Moreover, the actual raw sleep study data can be obtained among patients with an OSA diagnosis, thereby permitting additional analytical strategies. For example, quantitative endpoints can be extracted from sleep study data to conduct genetic association analyses of disease severity measures that complement association testing with EHR-based case/noncase status.

Case identification criteria for an EHR-based algorithm serve a different purpose than OSA screening tools, of which a number currently exist.^{15,23–25} Screening tools are developed to determine the likelihood of disease and help stratify patients into risk groups for clinical decisions about testing for sleep apnea and are useful for a variety of other purposes. In research, the probability of having a diagnosed condition can be used in combination with a case identification algorithm such as the one proposed here to adjust analyses of case-control status, to match diagnosed cases to similar but as-yet undiagnosed noncases, or to select low-risk patients among all potential noncases (thereby minimizing the proportion of undiagnosed disease). A limitation to using screening tools is that useful symptom data may be difficult to extract from the medical record on a large scale and may be unavailable in patients not already suspected to have the disease of interest.

Ultimately, the proposed EHR-based algorithm showed strong performance at each site, albeit with some variation. If a specific target performance characteristic is desired, it is advisable to perform a validation study in each unique source population. For example, the addition of a single instance of a diagnosis code at Geisinger increased the PPV from 87.5 to 94.9%. However, this 7% to 8% improvement in PPV came at the cost of a nearly 20%

Figure 3—Site-specific sensitivity and specificity of primary EHR-based OSA identification algorithm.

Sensitivity and specificity of the EHR-based OSA case definition of two or more ICD-9/ICD-10 diagnostic codes related to OSA is shown overall and at each participating site as a function of PPV, NPV and assumed prevalence. EHR = electronic health record, ICD = International Classification of Diseases, KPSC = Kaiser Permanente Southern California, NPV = negative predictive value, OSA = obstructive sleep apnea, Penn = University of Pennsylvania, PPV = positive predictive value.

reduction in the number of determinate cases. The reduction in the number of determinate cases was even more pronounced at the University of Pennsylvania, where 47% and 79% of EHR-defined cases became indeterminate when requiring at least three and at least four instances of diagnostic codes, respectively. This likely reflects the fact that the Hospital of the University of Pennsylvania has a large proportion of patients receiving tertiary care and procedures, which results in less opportunity for ongoing care and repeated coding for a given individual. Ultimately, requiring only two instances of diagnostic codes resulted in a high PPV and NPV in this validation study. However, site-specific differences highlight the importance of validating algorithm performance across different health systems, particularly those that may have unique characteristics not well captured at institutions in the current study. When choosing which criteria to implement, the relative value of potentially moderate improvements in PPV against large reductions in determinate cases should be carefully considered.

The benefit of using multiple OSA-related codes is supported by a recent study examining the validity of administrative data for identification of OSA prior to diagnostic testing.²⁶ Among patients referred for sleep testing, Laratta et al demonstrated that an algorithm of two outpatient billing claims or one hospital discharge code identifying OSA in the 2 years prior to testing had a PPV of 74.8%, but an NPV of only 18.3%, for detecting cases defined as a respiratory disturbance index ≥ 5 events/h.²⁶ The PPV of the algorithm was improved to 90.5% with the additional criteria of codes for OSA-related comorbidities (diabetes, hypertension, stroke, or acute myocardial infarction). Key differences between this prior study and the current analysis, which likely account for differences in performance

characteristics (particularly NPV), are the study population and more restricted timeframe. In particular, Laratta et al studied patients who had been referred for diagnostic testing between 2005 and 2007 and excluded patients with previous diagnosis or prior diagnostic testing. Thus, the sample was enriched for patients with undiagnosed OSA at higher risk of OSA. Conversely, our study includes a broad sample of patients from multiple health systems. In addition, the diagnostic algorithm used by Laratta et al focused on codes from the 2 years prior to diagnostic testing, rather than using the full medical history (as in the current study). Ultimately, although the analysis sample and timeframe used by Laratta et al were more limited in scope than the current study, the observation of high PPV when using multiple diagnostic codes supports the validity of the algorithm proposed here. In addition, data from Laratta et al²⁶ suggest that inclusion of diagnostic codes for OSA-related comorbidities (which are readily available in EHR-linked biorepositories) may present an alternative approach to improving PPV.

Although the described algorithm performs well with respect to identifying diagnosed cases, OSA is a progressive disorder that is often underdiagnosed and unrecognized.¹¹ The influence of undiagnosed OSA on case identification algorithms has been discussed extensively with respect to surgical patients.^{27,28} Using EHR data in the 2 years prior to surgery among patients referred for preoperative PSG, McIsaac et al concluded that the use of single diagnostic codes (or their combination) is inadequate for identifying patients with OSA.²⁸ Although this concern is valid, the magnitude of the problem may be overstated when focusing on surgical patients referred for PSG, as opposed to the general health system population (as in the current results). In particular, samples referred for testing are

likely at higher baseline risk for OSA and thus greatly enriched for patients with undiagnosed disease. As discussed, it is also likely the reliance on single occurrences of diagnostic codes, rather than multiple instances, reduces accuracy because codes are at times applied prematurely to justify diagnostic testing (ie, PSG). Similarly, leveraging the entire health record may increase accuracy compared to relying on a 2-year window prior to surgery or PSG. Ultimately, limitations caused by misclassification bias should be recognized and accounted for (eg, by controlling for known risk factors) and future analyses should be performed to understand the effect of undiagnosed disease in EHR-defined noncases using the current algorithm (eg, by performing sleep testing in a random subset). Although there are undoubtedly some noncases with or who will eventually present with OSA in our sample, almost 70% of EHR-defined noncases had predicted OSA probabilities below 20% based on the Symptomless Multivariable Apnea Prediction Score,¹⁵ suggesting a low likelihood of OSA in most noncases.

Our study has some limitations. As discussed, noncases are simply patients in whom OSA has not been diagnosed, and the NPV pertains to evidence of diagnosis rather than true absence of disease. Given that male sex is a risk factor for OSA, it is feasible that physicians are more likely to suspect sleep apnea in males, which could lead to a higher proportion of undiagnosed OSA among females.²⁹ Our study populations were drawn and validated independently at six different institutions. Thus, there may have been some differences in the assembly of the underlying validation samples. Data were generally enriched for or restricted to those who contributed a genetic sample at each site. These individuals may have been more educated, more involved in their own health care, or enriched for other disorders (eg, cardiovascular disease, cancer, and type II diabetes) if data were genotyped by investigators interested in research pertaining to specific conditions. However, the algorithm performed well at the University of Pennsylvania, which leveraged patients from the entire health system. Moreover, patients were required to have at least 1 year of enrollment/participation in the health system; this could exclude patients with frequent changes in health insurance or providers. Clinical chart reviews were necessarily performed by different individuals at each of the participating sites with access to details in the electronic medical records. Although common guidelines were used, this could lead to some variation in the approaches to validation of case or noncase status. However, our approach showed high reliability when comparing a random sample of clinical diagnoses between two independent reviewers at each site. Despite these potential differences, we were able to show similar and well-performing predictive characteristics at each site, further highlighting the utility of the proposed algorithm. Although our study sites had a diversity of source populations from which validation samples were drawn, results may not generalize as well to other types of health systems, such as less integrated health systems or safety-net systems that include more uninsured or vulnerable patients. Most participants included in this validation study were Caucasian and, thus, results may be less generalizable to other ethnic groups. Additional validation may be required to fully understand algorithm performance when

expanding the definition proposed here to any new institution, but in particular at those institutions known for unique patient characteristics.

CONCLUSIONS

In this study, we developed and validated a simple EHR-based algorithm for the identification of diagnosed cases of OSA, and conversely, undiagnosed noncases. The algorithm, which relies solely on instances of ICD-9/ICD-10 diagnostic codes, had excellent performance and was easy to apply across various institutions. A simple adjustment to the required number of code instances can be used to increase the PPV, albeit at the expense of the number of determinate cases. Although future research is needed to evaluate the effect of undiagnosed OSA among EHR-defined noncases, this case identification algorithm can be used to enhance a number of current research applications, including defining cases and noncases in existing genetic biorepositories for analyses of genetic associations. The algorithm may also be of use to other clinical investigators, clinicians, and population health care managers. Ultimately, the future applications of this simple case-identification algorithm are widespread, underscoring its potential utility.

ABBREVIATIONS

- AHI, apnea-hypopnea index
- BMI, body mass index
- CI, confidence interval
- CPAP, continuous positive airway pressure
- EHR, electronic health record
- ICD, International Classifications of Disease
- NPV, negative predictive value
- OSA, obstructive sleep apnea
- PAP, positive airway pressure
- PPV, positive predictive value
- PSG, polysomnography

REFERENCES

1. American Academy of Sleep Medicine. *International Classification of Sleep Disorders*. 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014.
2. Institute of Medicine. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. Washington, DC: The National Academies Press; 2006.
3. Loke YK, Brown JW, Kwok CS, Niruban A, Myint PK. Association of obstructive sleep apnea with risk of serious cardiovascular events: a systematic review and meta-analysis. *Circ Cardiovasc Qual Outcomes*. 2012;5(5):720–728.
4. Redline S, Yenokyan G, Gottlieb DJ, et al. Obstructive sleep apnea-hypopnea and incident stroke: the sleep heart health study. *Am J Respir Crit Care Med*. 2010;182(2):269–277.
5. Gozal D, Farre R, Nieto FJ. Obstructive sleep apnea and cancer: Epidemiologic links and theoretical biological constructs. *Sleep Med Rev*. 2016;27:43–55.
6. Olaithé M, Bucks RS, Hillman DR, Eastwood PR. Cognitive deficits in obstructive sleep apnea: insights from a meta-review and comparison with deficits observed in COPD, insomnia, and sleep deprivation. *Sleep Med Rev*. 2018;38:39–49.

7. Tregear S, Reston J, Schoelles K, Phillips B. Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis. *J Clin Sleep Med*. 2009;5(6):573–581.
8. Garbarino S, Guglielmi O, Sanna A, Mancardi GL, Magnavita N. Risk of occupational accidents in workers with obstructive sleep apnea: systematic review and meta-analysis. *Sleep*. 2016;39(6):1211–1218.
9. Baldwin CM, Griffith KA, Nieto FJ, O'Connor GT, Walsleben JA, Redline S. The association of sleep-disordered breathing and sleep symptoms with quality of life in the sleep heart health study. *Sleep*. 2001;24(1):96–105.
10. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res*. 2005;40(5 Pt 2):1620–1639.
11. Punjabi NM. The epidemiology of adult obstructive sleep apnea. *Proc Am Thorac Soc*. 2008;5(2):136–143.
12. Newton KM, Peissig PL, Kho AN, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Inform Assoc*. 2013;20(e1):e147–e154.
13. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–381.
14. Kelly H, Bull A, Russo P, McBryde ES. Estimating sensitivity and specificity from positive predictive value, negative predictive value and prevalence: application to surveillance systems for hospital-acquired infections. *J Hosp Infect*. 2008;69(2):164–168.
15. Lyons MM, Keenan BT, Li J, et al. Symptomless multi-variable apnea prediction index assesses obstructive sleep apnea risk and adverse outcomes in elective surgery. *Sleep*. 2017;40(3).
16. Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. Increased prevalence of sleep-disordered breathing in adults. *Am J Epidemiol*. 2013;177(9):1006–1014.
17. Schwab RJ, Badr SM, Epstein LJ, et al. An official American Thoracic Society statement: continuous positive airway pressure adherence tracking systems. The optimal monitoring strategies and outcome measures in adults. *Am J Respir Crit Care Med*. 2013;188(5):613–620.
18. Hwang D. Monitoring progress and adherence with positive airway pressure therapy for obstructive sleep apnea: the roles of telemedicine and mobile health applications. *Sleep Med Clin*. 2016;11(2):161–171.
19. Budhiraja R, Thomas R, Kim M, Redline S. The role of big data in the management of sleep-disordered breathing. *Sleep Med Clin*. 2016;11(2):241–255.
20. Pépin JL, Tamisier R, Hwang D, Mereddy S, Parthasarathy S. Does remote monitoring change OSA management and CPAP adherence? *Respirology*. 2017;22(8):1508–1517.
21. Emamian F, Khazaie H, Tahmasian M, et al. The association between obstructive sleep apnea and Alzheimer's disease: a meta-analysis perspective. *Front Aging Neurosci*. 2016;8:78.
22. Yaffe K, Laffan AM, Harrison SL, et al. Sleep-disordered breathing, hypoxia, and risk of mild cognitive impairment and dementia in older women. *JAMA*. 2011;306:613–619.
23. Maislin G, Pack AI, Kribbs NB, et al. A survey screen for prediction of apnea. *Sleep*. 1995;18(3):158–166.
24. Chung F, Subramanyam R, Liao P, Sasaki E, Shapiro C, Sun Y. High STOP-Bang score indicates a high probability of obstructive sleep apnoea. *Br J Anaesth*. 2012;108(5):768–775.
25. Jonas DE, Amick HR, Feltner C, et al. Screening for obstructive sleep apnea in adults: evidence report and systematic review for the US Preventive Services Task Force. *JAMA*. 2017;317(4):415–433.
26. Laratta CR, Tsai WH, Wick J, Pendharkar SR, Johannson KA, Ronksley PE. Validity of administrative data for identification of obstructive sleep apnea. *J Sleep Res*. 2017;26(2):132–138.
27. Finkel KJ, Searleman AC, Tymkew H, et al. Prevalence of undiagnosed obstructive sleep apnea among adult surgical patients in an academic medical center. *Sleep Med*. 2009;10(7):753–758.
28. McIsaac DI, Gershon A, Wijeyesundara D, Bryson GL, Badner N, van Walraven C. Identifying obstructive sleep apnea in administrative data: a study of diagnostic accuracy. *Anesthesiology*. 2015;123(2):253–263.
29. Wimms A, Woehrle H, Ketheeswaran S, Ramanan D, Armitstead J. Obstructive sleep apnea in women: specific issues and interventions. *BioMed Res Int*. 2016;2016:1764837.

ACKNOWLEDGMENTS

The authors thank Dr. Diego Mazzotti at the University of Pennsylvania for contributions and helpful discussion regarding the sample at Penn.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication June 10, 2019

Submitted in final revised form September 13, 2019

Accepted for publication September 13, 2019

Address correspondence to: Brendan T. Keenan, 125 South 31st St., Suite 2100, Office 2117, Philadelphia, PA 19104, Tel: (215) 746-6812; Fax: (215) 746-4814; Email: keenanbr@pennmedicine.upenn.edu

DISCLOSURE STATEMENT

All authors have seen and approved of the final manuscript. Work for this study was performed at the University of Pennsylvania, Geisinger, Mayo Clinic, Northwestern University, Kaiser Permanente Southern California, and Vanderbilt University Medical Center. The research presented here was supported by the National Institutes of Health Grants R01 HL134015 (Approaches to Genetic Heterogeneity of Obstructive Sleep Apnea) and P01 HL094307 (Individual Differences in Obstructive Sleep Apnea). This project was also funded, in part, under a grant with the Pennsylvania Department of Health (#SAP 4100070267). Dr. Pack is The John L. Miclot Professor of Medicine at the University of Pennsylvania; funds for this endowment were provided by the Philips Respironics Foundation. Dr. Morgenthaler serves as a consultant to Respicardia, Inc. All other authors report no conflicts of interest.