Prevalence and Characterization of Yoga Mentions in the Electronic Health Record

Nadia M. Penrod, PhD, Selah Lynch, MS, Sunil Thomas, MBA, Nithya Seshadri, BS, and Jason H. Moore, PhD

Background: **There is a growing patient population using yoga as a therapeutic intervention, but little is known about how yoga interfaces with health care in clinical settings.**

Purpose: **To characterize how yoga is documented at a large academic medical center and to systematically identify clinician-derived therapeutic use cases of yoga.**

Methods: **We designed a retrospective observational study using a yoga cohort (n = 30,976) and a demographically matched control cohort (n = 92,919) from the electronic health records at Penn Medicine between 2006 and 2016. We modeled the distribution of yoga notes among patients, clinicians, and clinical service departments, built a multinomial Naïve Bayes classifier to separate the notes by context-dependent use of the word yoga, and modeled associations between clinician recommendations to use yoga and 754 diagnostic codes with Fisher's exact test, setting an false discovery rate (FDR)-adjusted *P*-value ≤ .05 (ie, *q*-value) as the significance threshold.**

Results: **Yoga mentions in the electronic health record have increased 10.4-fold during the 10-year study period, with 2.6% of patients having at least 1 mention of yoga in their notes. In total, 30,976 patients, 2398 clinicians, and 41 clinical service departments were affiliated with yoga notes. The majority of yoga notes are in primary care. Nine diagnoses met the significance criteria for having an association with clinician recommendations to use yoga including Parkinson's disease (Odds ratio [OR], 6.3 [3.7 to 11.4]; *q*-value < 0.001), anxiety (OR, 5.8 [3.8 to 9.0]; *q*-value < 0.001), and backache (OR, 3.8 [2.4 to 6.3]; *q*-value = 0.001).**

Conclusions: **There is a widespread and growing trend to include yoga as part of the clinical record. In practice, clinicians are recommending yoga as a nonpharmacological intervention for a subset of common chronic diseases. (J Am Board Fam Med 2019;32:790–800.)**

Keywords: **Academic Medical Centers, Behavioral Medicine, Chronic Disease Management, Electronic Health Records, Life Style, Medical Informatics, Meditation, Primary Health Care, Yoga**

Patients are turning to complementary therapies (eg, yoga, meditation, Mindfulness-Based Stress Reduction, etc.) in increasing numbers in an effort to prevent or treat disease and to manage symptoms, drug side-effects, and chronic stress.[1,2] The use of complementary therapies in addition to the standard of care is known as integrative medicine, a field that continues to gain traction despite the fact that most published research is constrained by small sample sizes, inadequate protocol descriptions, and the absence of quantitative or objective outcome measures.

Studies in this domain present a challenging research design problem. They are not amenable to customary double-blind placebo-controlled trials,

Corresponding author: Jason H. Moore, PhD, D202 Richards Building, 3700 Hamilton Walk, University of Pennsylvania, Philadelphia, PA 19104-6116 (E-mail: jhmoore@upenn.edu).

they are conceptually and practically difficult to dose, and the effects are highly subjective.[3] These challenges spotlight the need for data that captures how complementary therapies interface with health care, not in a controlled setting, but in the real world.

A potentially abundant source of such data is the electronic health record (EHR). Leveraging social and behavioral determinants of health from EHR data has promising implications for population-based research and public health.[4–6] As it stands, social and behavioral factors are captured in the clinical chart notes. Chart notes mentioning integrative medicine represent the practical experiences of patients and providers as a type of expert knowledge that does not rely on recall which can bias survey data. Through text-mining and natural language processing, this information becomes a valuable source of empirical data to begin filling evidence gaps and to establish effective use cases for integrative medicine.

Yoga, a practice based on controlled breathing, movement, and meditation, is the most commonly used type of integrative medicine. An estimated 1 in 7 American adults used yoga in 2017, according to the Centers for Disease Control and Prevention.[2] When surveyed, 81.9% of yoga practitioners reported overall improvements in physical and mental health.[7] Moreover, systematic reviews suggest yoga promotes healthy aging and may be an effective treatment strategy for patients that have chronic diseases with stress or anxiety-related co-morbidities.[8–10]

Here, we present a retrospective observational study in the EHR that characterizes how yoga is documented in our health care system, and identifies clinician-derived therapeutic use cases of yoga.

Methods

Data Collection

We identified study and control cohorts by querying the EHR at the University of Pennsylvania Health System. First, we used PennSeek, a tool that aids unstructured text search in the EHR, to identify patients with the word "yoga" in their outpatient chart notes between November 15, 2006 and November 16, 2016. Then, using the unique identifiers for each of these patients, we queried the Penn Data Store, a database housing structured EHR data, to retrieve associated demographic and diagnostic information. We combined the structured and unstructured data to generate a yoga cohort of 30,976 patients.

We also identified a set of demographically matched controls from the Penn Data Store by randomly sampling outpatient records to match the yoga cohort based on age, sex, and race. We used an approximate 3:1 control-to-case ratio to increase statistical power and coverage across unknown or unmeasured confounders.[11] The control cohort contains 92,919 patients.

For each data set, the inclusion criteria were at least 1 encounter from an outpatient office or allied health visit between November 15, 2006 and November 16, 2016, at least 1 primary diagnostic code, and age ≥ 19 years. All patients are represented in the data set by a single (most recent) encounter. For the yoga cohort, this is the most recent encounter with a chart note mentioning yoga. To capture a snapshot of the entire Penn Medicine patient population during the study period, we also collected demographic data on all patients meeting the inclusion criteria above (n = 1,210,228).

All data were accessed through the Data Analytics Center at Penn Medicine and the study protocol was approved by the Institutional Review Board (IRB) at the University of Pennsylvania.

Demographics

To report demographics (age, sex, race, ethnicity, and financial class) we used data from the most recent encounter for each patient. Statistical comparisons between groups were calculated by the t-test for the continuous variable, age, and the χ^2 test for the categorical variables: sex, race, ethnicity, and financial class.

Diagnostic Codes

Medical conditions were recorded with the International classification of diseases versions 9 and 10 (ICD-9 and ICD-10). We first used the ICD-9-CM and ICD-10-CM general equivalence maps from the AHRQ MapIT Software (Rockville, MD) to create a single unified map from ICD code to medical condition and then we collapsed disease name synonyms with the National Center for Biomedical Ontology (NCBO) human disease ontology.[12–14]

The medical conditions associated with an encounter are filtered to include only the ICD codes given a primary diagnosis designation.

Classification

We built a text classifier to label the context of the word yoga in each chart note.

Feature Set

To generate text features that can be used to predict class, we applied a standard natural language-processing pipeline to the chart notes. First, we preprocessed the text to remove symbols, punctuation, and case before breaking down the paragraphs into individual words. The tense of the words was not modified by stemming or lemmatization and stop words (eg, the, and, or, etc.) were retained. Second, to represent each note, we selected the word "yoga" and a given number words on either side of yoga to generate a context-window. We used context-windows to reduce noise because most records have a single mention of the word yoga among hundreds of words of text unrelated to yoga. Third, from the context-windows, we generated ngrams of length 1, 2, and 3, and skipgrams of length 2. Here, ngrams are contiguous sequences of words and skipgrams are constructed by pairing the first and third word in a contiguous sequence. We required each ngram and skipgram to appear in at least 2 chart notes, and excluded those that appeared in greater than 95% of chart notes. The ngrams and skipgrams from all the chart notes were combined to create a single vocabulary. Text processing was done in Python version 3.6.1 with the Natural Language Toolkit (NLTK version 3.2.4), a free, open-source platform for analyzing human language data.[15]

Classifier

To categorize the chart notes based on the context of the word yoga, we built a multinomial Naïve Bayes classifier. Multinomial Naïve Bayes is a supervised classifier used to predict the probability of a note belonging to a given class, based on its features (ie, ngrams and skipgrams). These models can handle large sets of features including irrelevant features and features with equal predictive probabilities, and they are fast. For our text classification problem, the multinomial Naïve Bayes model outperformed support vector machines, random forests, and gradient-boosting models.

The model was trained, tuned, and tested on a subset of manually annotated chart notes (n = 5600). In the annotation process, we identified 3 classes: clinician-documented yoga, clinician-rec-

ommended yoga, and other miscellaneous mentions of the word yoga. Each note is assigned 1 of these 3 class labels. The model is trained to identify patterns between the features and the class labels by iteratively evaluating subsets of the annotated data. To address class imbalance, we used stratified 5-fold cross validation to train the model and to optimize hyperparameters for feature selection and modeling. Eighty percent of the annotated notes were used to train the model (training set), the other 20% were used to test the performance of the model (test set). To evaluate the performance of the model, we calculated balanced accuracy in the test set. In this case, balanced accuracy is the weighted average of the sensitivity score, that is, the true positive rate, for each class. The final model uses mutual information for feature selection, retains features with scores in the top 50% percentile, and assigns the α parameter in the multinomial Naïve Bayes model to 1.0, which is incidentally the default. Classification was done in Python version 3.6.1 with Scikit-learn (sklearn version 0.18.1),[16] a free, open-source software library for machine learning.

Statistical Analyses

To characterize trends over time in the numbers of chart notes mentioning yoga, and the corresponding numbers of patients, clinicians, and clinical service departments, we fit linear and quadratic models to the unique number of yoga notes by year. Quadratic models were selected because they minimized the Akaike Information Criterion.[17] To characterize trends over time in the adoption of yoga as part of the clinical record by clinical service department, we fit linear models to the unique number of yoga notes recorded in a given department by year.

To identify medical conditions associated with clinician recommendations to use yoga, we calculated the odds of having a given ICD code and a yoga recommendation at the same clinical encounter. To do this, we used Fisher's exact test to calculate odds ratios and P-values based on the counts of patients with a given ICD code in the clinician-recommended yoga class relative to those of their demographically matched controls. To correct for multiple testing, the significance threshold for inclusion was an false discovery rate (FDR)-adjusted P-value (ie, q-value) of ≤ 0.05. To ensure we did not select for special cases, we filtered the statisti-

cally significant results to include only medical conditions for which there were at least 25 patients and 5 clinicians in the yoga class.

Statistical analyses were performed with R version 3.5.1 and plots were generated with the R package ggplot2 version 3.0.0.[18,19]

Results

Prevalence of Yoga in the EHR

The yoga data set was generated from a keyword search for "yoga" in the outpatient clinical chart notes. Through this search we identified 61,976 unique notes with yoga mentions corresponding to a cohort of 30,976 unique patients (median, 1 note per patient; interquartile range, 1 to 2 note(s) per patient). The yoga cohort accounted for 2.6% of the total patient population in the database during the study period. A control cohort was selected to match the yoga cohort on age, sex, and race by randomly sampling outpatient records that did not mention yoga in a 3:1 control to case ratio (n = 92,919).

The yoga cohort is enriched for younger, non-Hispanic, white females with commercial insurance relative to the broader patient population at Penn Medicine (Table 1).

Documentation of yoga in the EHR at Penn Medicine is trending upward. From November 2006 to November 2016, there was a 10.6-fold increase in the number of clinical chart notes mentioning the word yoga, an 8.9-fold increase in the number of patients with 1 or more notes mentioning yoga, a 5.4-fold increase in the number of clinicians using chart notes to document yoga, and a 1.2-fold increase in the number of clinical service departments affiliated with notes mentioning yoga. In all cases, these increases showed quadratic growth (Figure 1).

In total, 2398 unique clinicians documented yoga in the EHR during the study period. To ensure we have not selected for a subset of clinicians that have a penchant for yoga, we verified that 99.3% (2,381) of these clinicians are included as providers in the control cohort.

Together, the yoga and control cohorts come from a total of 55 unique clinical service departments. Forty-one of these departments (75%) had at least 1 note mentioning yoga on record during the study period. All clinical service departments represented in the yoga cohort are also represented in the control cohort.

The distribution of notes mentioning yoga, and the time to adoption of yoga as part of the clinical record, vary across clinical service departments (Supplemental Figure 1). The largest numbers of yoga notes were found in the primary care setting. In 2016, primary care alone accounted for 4904 yoga notes, or 38.2% of the total number of notes mentioning yoga that year. Linear models fit to count data illustrate that once the word yoga appears in a department, its use tends to grow (Supplemental Figure 1).

Characterization of Yoga in the EHR

To determine the context of the word yoga in the chart notes, we built a supervised text classification pipeline based on a set of 5600 manually annotated notes. In the annotation process, we identified 3 classes: clinician-documented yoga, clinician-recommended yoga, and other miscellaneous mentions of yoga.

The clinician-documented yoga class includes cases where the clinician is recording a patient's self-reported yoga practice, either as a lifestyle, for example, "exercise activity: yoga, 60 minutes/day, 3 day/week," or as a therapeutic intervention, for example, "anxiety – managed with yoga only and no medications." The clinician-recommended yoga class includes cases where the clinician is recommending yoga as a therapeutic intervention, for example, "I have recommended that she return to see her physiatrist, and also consider other forms of therapy including yoga." And the other miscellaneous mentions of yoga class includes cases of ambiguous semantics, references to commercial products, like yoga pants or yoga toes (YogaToes, Dexter, MI), a product advertised to relieve foot pain, and automatically generated contraindications, for example, "avoid high-velocity sports (downhill skiing); avoid head stands and plow [pose] in yoga—use of high head rest and seatbelt/chest belt while in a car."

The annotated notes were used to build a supervised multinomial Naïve Bayes classifier (balanced accuracy = 0.87, Supplemental Figure 2, see Methods). The tuned model used features generated from context windows of length 11, that is, the word "yoga" ± 5 words, and the final feature set contained 8508 elements. Discriminating features included the following: meditation, try, discussed, week, consider, does, stress, advised, doing, breathing, plow in yoga, head stands, tried yoga, exercise, therapy, stretching. The classifier was used to as-

Table 1. Demographic Characteristics

Characteristic	Patients, No (%)			P-Value[†] (Control, Penn Med)
	Yoga Cohort (n = 30,976)	Control Cohort (n = 92,919)	Health System (n = 1,210,228)	
Age, y				
19 to 44	14,069 (45.4)	42,158 (45.4)	478,886 (39.6)	
45 to 64	11,843 (38.2)	35,554 (38.3)	425,661 (35.2)	
65 to 84	4838 (15.6)	14,535 (15.6)	266,283 (22.0)	
85+	226 (0.7)	672 (0.7)	39,398 (3.3)	
Median age (IQR), y	47 (34 to 60)	47 (34 to 60)	51 (34 to 65)	P = .996, < .001
Sex				
Female	25,688 (82.9)	77,063 (82.9)	721,180 (59.6)	
Male	5288 (17.1)	15,856 (17.1)	489,048 (40.4)	P = .984, < .001
Race				
White	23,259 (75.1)	69,777 (75.1)	785,163 (64.9)	
Black or African American	3625 (11.7)	10,874 (11.7)	243,639 (20.1)	
Other*	2744 (8.9)	8231 (8.9)	135,081 (11.2)	
Asian	1289 (4.2)	3867 (4.2)	43,514 (3.6)	
American Indian or Alaskan Native	35 (0.1)	98 (0.1)	1155 (0.1)	
Native Hawaiian or Other Pacific Islander	24 (0.1)	72 (0.1)	1676 (0.1)	P = 1.00, < .001
Ethnicity				
Non-Hispanic	30,189 (97.5)	89,272 (96.1)	1,167,368 (96.5)	
White, Hispanic	513 (1.7)	2913 (3.1)	28,732 (2.4)	
Black or African American, Hispanic	128 (0.4)	338 (0.4)	8414 (0.7)	
Other*, Hispanic	74 (0.2)	263 (0.3)	3933 (0.3)	
Asian, Hispanic	70 (0.2)	129 (0.1)	1658 (0.1)	
American Indian or Alaskan Native, Hispanic	1 (0.0)	4 (0.0)	31 (0.0)	
Native Hawaiian or Other Pacific Islander, Hispanic	1 (0.0)	0 (0.0)	92 (0.0)	P < .001, < .001
Financial				
Commercial	19,978 (64.5)	54,634 (58.8)	619,077 (51.2)	
Not recorded	5826 (18.8)	18,770 (20.2)	246,673 (20.4)	
Medicare	4127 (13.3)	13,246 (14.3)	254,220 (21.0)	
Medicaid	1042 (3.4)	6229 (6.7)	89,469 (7.4)	
Self-pay	3 (0.0)	40 (0.0)	789 (0.1)	P < .001, < .001

IQR, interquartile range.

*Other includes mixed race, other, and unknown.

[†]Variables were compared between groups using a *t*-test for the continuous variable age, and the χ^2 test for the categorical variables: sex, race, ethnicity, and financial class. Comparisons were made between the yoga cohort and the control cohort (*P*-values before the comma) and between the yoga cohort and the broader patient population (*P*-values after the comma).

sign each yoga note to 1 of the 3 predefined classes. Based on this classification, clinician-documented yoga accounted for approximately 75% of yoga mentions, other miscellaneous mentions for approximately 14% of yoga mentions, and clinician-recommended yoga for approximately 10% of yoga mentions in the chart notes.

To determine if the distributions of yoga notes among the 3 classes changed over time, we looked at the distribution of yoga notes by class and year. As the total number of yoga notes increased each year, the numbers of notes in each class also increased (Figure 2A). But the fraction of the total number of yoga notes represented by each class remained relatively constant (Figure 2B). In other words, the number of yoga notes in each class increased in proportion to the total number of yoga notes written each year. On

Figure 1. Documentation of yoga in the electronic health record (EHR) shows quadratic growth. Quadratic models fit to count data for the number of unique yoga note(s) per year, the number of unique patients with yoga notes, the number of unique clinicians writing yoga notes, and the number of unique clinical service departments containing yoga notes. Quadratic models were selected over linear models based on the Akaike Information Criterion.
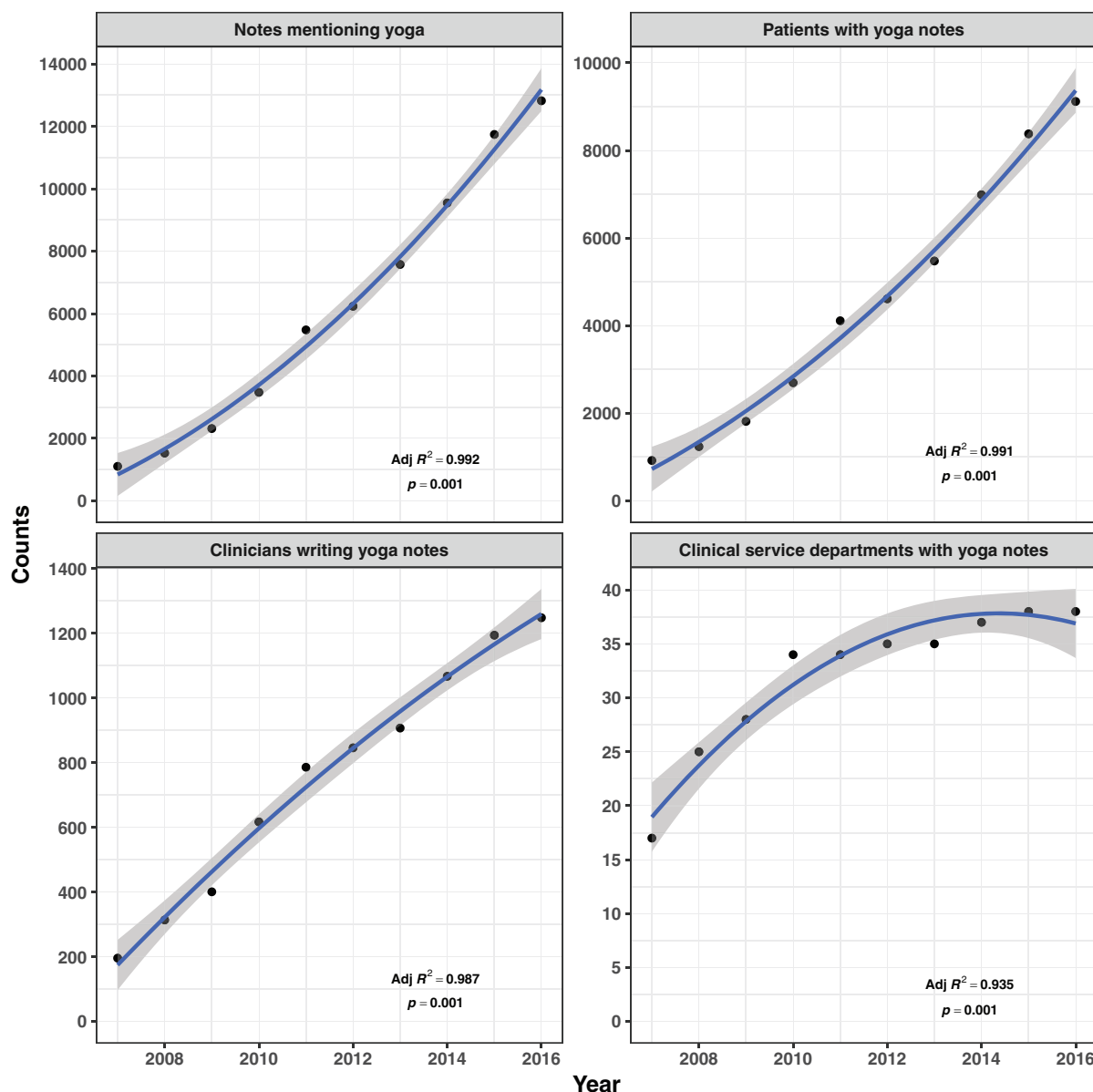
average, the distribution of yoga notes among the 3 classes did not change during the study period.

We focus the remaining analysis on the clinician-recommended yoga class.

EHR-Derived Use Cases of Yoga as Therapy

To systematically explore which medical conditions were associated with clinician recommendations to practice yoga, we used Fisher's exact test to compare the proportions of patients assigned each ICD code in the clinician-recommended yoga class to that of their demographically matched controls. Each code was tested independently, we did not combine similar or related ICD codes. Our threshold for significance was a predefined q-value ≤ 0.05.

Yoga mentions in the clinician-recommended yoga class were associated with 9 medical conditions (Figure 3). We consider these medical condi-

Figure 2. Clinician-documented yoga notes and clinician-recommended yoga notes increase in proportion to the total number of notes mentioning yoga each year. Bar plots show (A) counts of unique clinical chart notes mentioning yoga by class each year and (B) the proportion of clinical chart notes mentioning yoga by class each year.
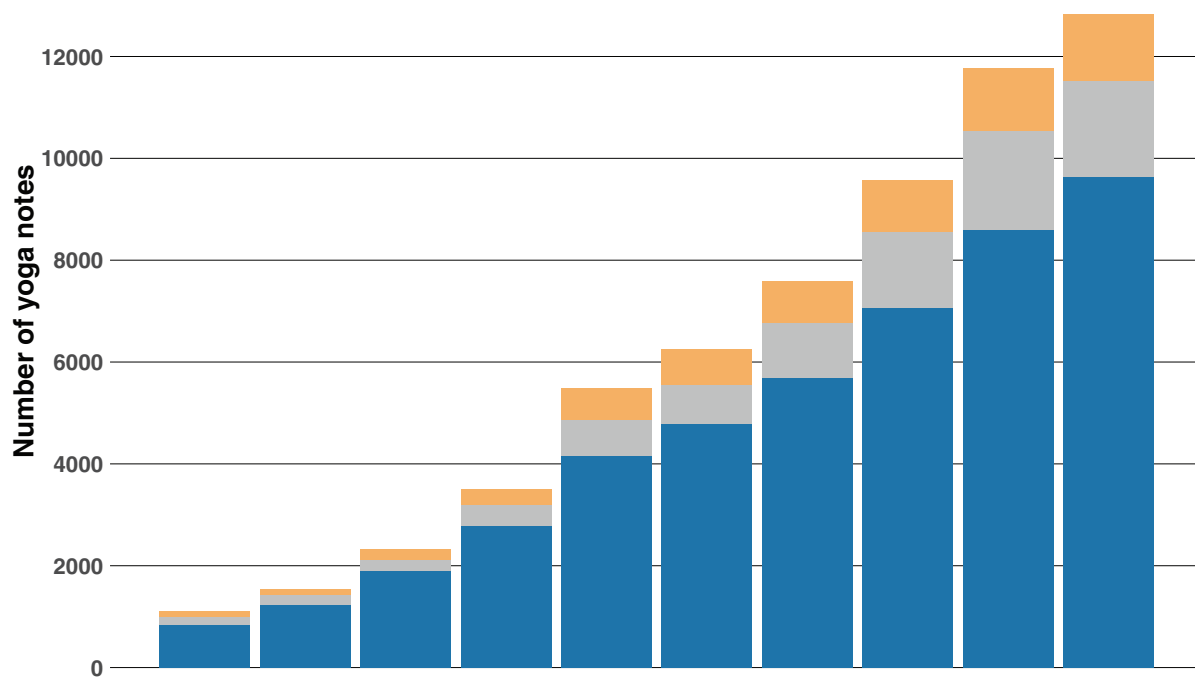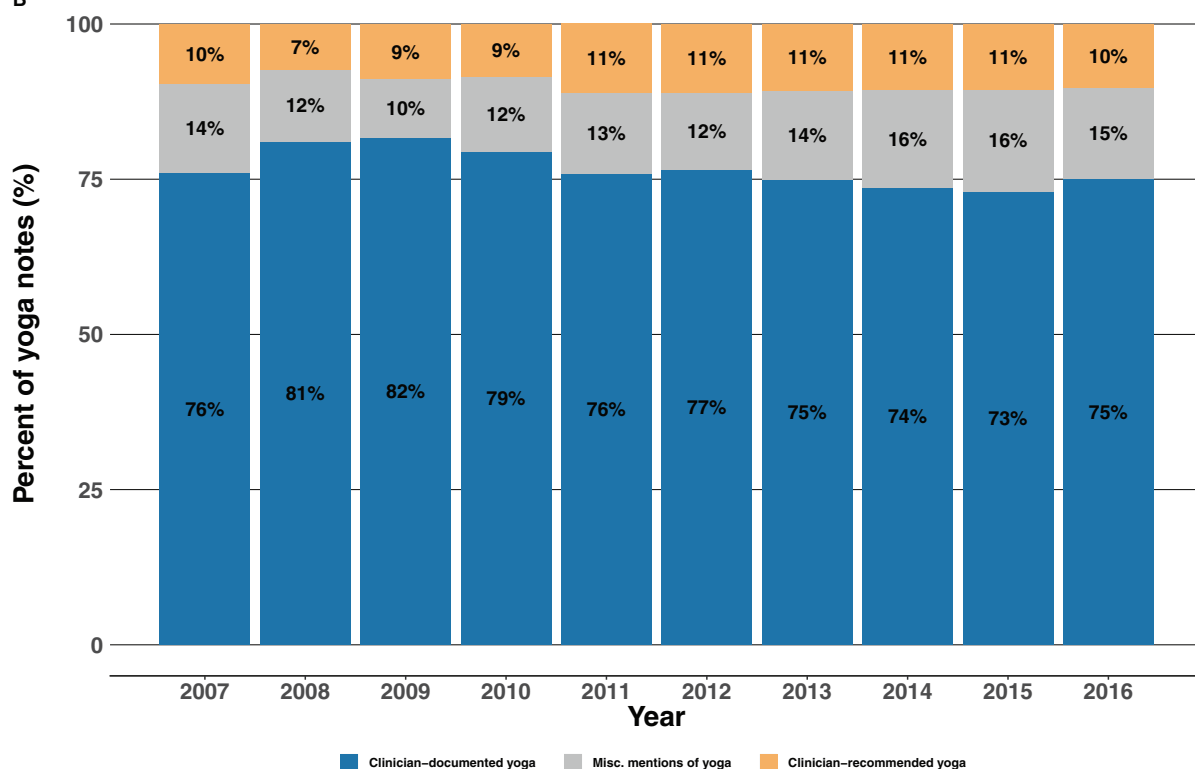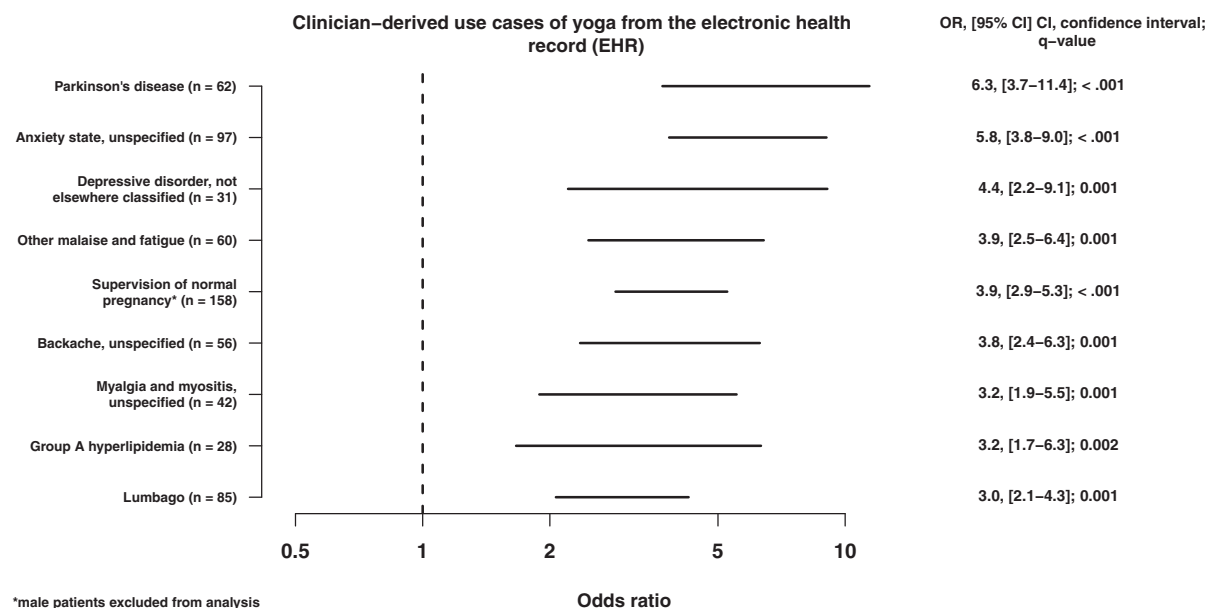
Figure 3. Clinician-recommended yoga notes are associated with 9 medical conditions. Odds ratios (OR) and P-values were calculated with Fisher's exact test using the counts of patients in the clinician-recommended yoga class for each medical condition (based on International Classification of Diseases (ICD) codes) and their controls, patients without a chart note mentioning yoga matched on age, sex, and race. Significance threshold for inclusion was a false discovery rate (FDR)-adjusted P-value ≤ .05.

Clinician–derived use cases of yoga from the electronic health record (EHR)

OR, [95% CI] CI, confidence interval; q–value

Parkinson's disease (n = 62) — 6.3, [3.7–11.4]; < .001

Anxiety state, unspecified (n = 97) — 5.8, [3.8–9.0]; < .001

Depressive disorder, not elsewhere classified (n = 31) — 4.4, [2.2–9.1]; 0.001

Other malaise and fatigue (n = 60) — 3.9, [2.5–6.4]; 0.001

Supervision of normal pregnancy* (n = 158) — 3.9, [2.9–5.3]; < .001

Backache, unspecified (n = 56) — 3.8, [2.4–6.3]; 0.001

Myalgia and myositis, unspecified (n = 42) — 3.2, [1.9–5.5]; 0.001

Group A hyperlipidemia (n = 28) — 3.2, [1.7–6.3]; 0.002

Lumbago (n = 85) — 3.0, [2.1–4.3]; 0.001

Odds ratio

*male patients excluded from analysis

tions to be clinician-derived therapeutic use cases of yoga. Parkinson's disease had the highest odds ratio (OR, 6.3; 95% CI, 3.7–11.4; q-value < 0.001). Other identified medical conditions meeting the significance criteria included: anxiety (OR, 5.8; 95% CI, 3.8–9.0; q-value < 0.001), depressive disorder (OR, 4.4; 95% CI, 2.2–9.1; q-value = 0.001), malaise and fatigue (OR, 3.9; 95% CI, 2.5–6.4; q-value = 0.001), pregnancy (OR, 3.9; 95% CI, 2.9–5.3; q-value < 0.001), backache (OR, 3.8; 95% CI, 2.4–6.3; q-value = 0.001), myalgia and myositis (OR, 3.2; 95% CI, 1.9–5.5; q-value = 0.001), hyperlipidemia (OR, 3.2; 95% CI, 1.7–6.3; q-value = 0.002), and lumbago (OR, 3.0; 95% CI, 2.1–4.3; q-value = 0.001).

Clinician recommendations to practice yoga came predominantly from primary care providers, with the exceptions of Parkinson's disease, pregnancy, and myalgia and myositis, which were most frequently found in notes attributed to providers from neurology, obstetrics and gynecology, and rheumatology departments, respectively.

Discussion

EHRs are an abundant data source that can uniquely fill the evidence gaps for some uses of complementary and integrative medicine in the prevention and treatment of chronic diseases. Here, we used EHR data from an academic medical center to characterize how yoga interfaces with health care.

We have shown that a large number of patients have medical chart notes that mention the word, yoga. Demographically, this subset of patients tracks national survey data showing that non-Hispanic, white, women are more likely to report using yoga in the past 12 months than men or individuals self-identifying as Hispanic, Black, or African American.[2] An under-representation of racial and/or ethnic minorities and individuals of low socioeconomic status engaging with yoga has been documented in the literature and reported barriers to use of yoga include preconceived ideas of the practice, family and work obligations, and yoga class location.[20] Increases in the number of patients with yoga notes are also supported by national trends showing use of yoga among US adults has been on the rise since 2002.[1,2]

In 2002, due to increased interest in and use of complementary and alternative medicine by patients and providers, the Federation of State Med-

ical Boards published guidelines around its use.[21] While there are no large-scale studies surveying health care provider perspectives on the role of yoga in medicine over time, our data show that each year an increasing number of clinicians, across clinical service departments, are at the minimum acknowledging yoga by documenting it in the chart notes of their patients.

There are reports suggesting that clinicians see potential in complementary and integrative medicine but do not feel they have enough information or experience to counsel patients on its proper use.[22] In some cases, clinicians are cautiously biding time until better research emerges.[23] This may explain why, despite the widespread and growing documentation of yoga in the EHR, clinician recommendations to use yoga remain proportionally unchanged.

Nevertheless, the EHR provides a unique data source to explore how yoga is currently being used in health care. From these data, we were able to identify 9 medical conditions that met the significance criteria for having an association with clinician recommendations to use yoga in practice.

Among these conditions, 2 are musculoskeletal: backache and lumbago. This finding is consistent with the American College of Physicians clinical practice guidelines that recommend yoga as a non-pharmacological, first-line therapy to treat chronic low back pain.[24]

We also identified 2 mental health conditions: anxiety and depressive disorder. There is an evidence-base supporting the use of yoga for these conditions.[25–27] One compelling aspect of this result is that these conditions are often comorbidities and the yoga literature is replete with studies of yoga as treatment for anxiety and depression in the context of other chronic diseases, including some of the conditions identified in this study.

The research on Parkinson's disease has explored the effects of yoga on the physical manifestations of the disease including motor control, postural stability, and functional mobility, and in the context of depression, fatigue and apathy, some of the most common comorbidities of the disease.[28–31]

In pregnancy, yoga research has focused on everything from safety for mothers and babies to mitigating low back pain, labor pain, anxiety, depression, stress, and sleep disturbances.[24–27]

We identified the endocrine and metabolic condition, hyperlipidemia. Meta-analyses of yoga for prevention of cardiovascular disease and stress-related biomarkers indicate there is suggestive but not conclusive evidence that yoga can effectively modulate lipid levels.[35,36]

Finally, we see evidence that the nonspecific symptoms of malaise and fatigue, and myalgia and myositis are being treated with yoga. This fits the narrative that patients prefer unconventional medicine, like diet-based or mind-body interventions, in cases of diagnostic uncertainty where there is ambiguous cause and effect.[37] For conditions that manifest these symptoms but do not have an effective standard of care, like fibromyalgia, yoga can be a beneficial component of a multipronged approach to symptom management.[38]

The medical conditions identified in this study demonstrate wide-ranging use cases of yoga as therapy and support the notion that the integrated systemic effects of yoga lead to perceived improvements in both physical and mental health. As our health care system transitions to a value-based, patient-centered model, the broader implication of this study is that yoga, as a multipurpose nonpharmacological intervention, has a place in the health care landscape.[39]

Limitations and Future Work

The challenges inherent to secondary use of EHR data for clinical research have been explored and described elsewhere.[40–43] In this work, we assume bias in patient reporting and clinician-documentation of yoga. There were no controls for the type/definition of yoga or for the frequency or duration of yoga practice. In addition, we relied on ICD codes as a proxy for diagnostic data as a means of maintaining consistency between the clinician-recommended yoga class and their demographically matched controls. It is possible that yoga was mentioned in a context not related to these codes. It is also possible that the odd ratios would change if we had combined ICD codes for similar or related conditions. ICD codes were not combined to maintain the integrity of the data as collected.

We acknowledge that with a balanced testing accuracy of 87%, there are some misclassifications in the clinician-recommended class adding noise to our analysis.

Future work will include in-depth longitudinal case studies in the EHR of patients with the iden-

tified medical conditions to assess changes in health outcomes attributable to yoga practice. As with other personalized medicines, we anticipate the effects of yoga to be patient dependent.

Conclusions

EHRs provide a unique opportunity to research yoga and other types of integrative medicine in a real-world setting. The prevalence of yoga mentions in the EHR is on the rise, and from these data we were able to use a kind of retrospective crowdsourcing based on the practical experience of clinicians to identify 9 therapeutic use cases of yoga. These results contribute to a growing body of evidence about the role integrative medicine plays in the treatment and management of chronic diseases. As our health care system evolves, and patients and clinicians turn to complementary therapies in greater numbers, we owe it to the community to continue to build this evidence-base by rigorously investigating the most effective use cases of these practices.

To see this article online, please go to: http://jabfm.org/content/ 32/6/790.full.

References

1. Clarke TC, Black LI, Stussman BJ, Barnes PM, Nahin RL. Trends in the use of complementary health approaches among adults: United States, 2002–2012. Natl Health Stat Rep 2015;10:1–16.

2. Clarke TC, Barnes PM, Black LI, Stussman BJ, Nahin RL. Use of yoga, meditation, and chiropractors among U.S. adults aged 18 and over. Hyattsville, MD: National Center for Health Statistics; 2018.

3. Davidson RJ, Kaszniak AW. Conceptual and methodological issues in research on mindfulness and meditation. Am Psychol 2015;70:581–592.

4. Adler NE, Stead WW. Patients in context—EHR capture of social and behavioral determinants of health. N Engl J Med 2015;372:698–701.

5. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. Annu Rev Public Health 2016;37:61–81.

6. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012;13:395–405.

7. Stussman BJ, Black LI, Barnes PM, Clarke TC, Nahin RL. Wellness-related use of common complementary health approaches among adults: United States, 2012. Hyattsville, MD: National Center for Health Statistics; 2015;12.

8. Sivaramakrishnan D, Fitzsimons C, Kelly P, et al. The effects of yoga compared to active and inactive controls on physical function and health related quality of life in older adults- systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2019;16:33.

9. Danhauer SC, Addington EL, Cohen L, et al. Yoga for symptom management in oncology: a review of the evidence base and future directions for research. Cancer 2019;125:1979–1989.

10. Desveaux L, Lee A, Goldstein R, Brooks D. Yoga in the management of chronic disease: a systematic review and meta-analysis. Med Care 2015;53:653–661.

11. Hennessy S, Bilker WB, Berlin JA, Strom BL. Factors influencing the optimal control-to-case ratio in matched case-control studies. Am J Epidemiol 1999;149:195–197.

12. Agency for Health Research and Quality. AHRQ MapIt Software Users Guide. 2018. Available from: https://www.qualityindicators.ahrq.gov/Downloads/ Resources/AHRQ_MapIT_FY2018_Installation_ Instructions.pdf. Accessed October 3, 2018.

13. Musen MA, Noy NF, Shah NH, et al. The national center for biomedical ontology. J Am Med Inform Assoc 2012;19:190–195.

14. Whetzel PL, Noy NF, Shah NH, et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res 2011;39(suppl):W541–W545.

15. Bird S, Looper E, Klein E. *Natural language processing with python*. Champaign, IL: O'Reilly Media Inc.; 2009.

16. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Machine Learning Res 2011;12:2825–2830.

17. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res 2004;33:261–304.

18. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from: https://www.R-project.org/.

19. Wickham H. Ggplot2: Elegant graphics for data analysis. New York, NY: Springer-Verlag; 2016. Available from: http://ggplot2.org.

20. Spadola CE, Rottapel R, Khandpur N, et al. Enhancing yoga participation: a qualitative investigation of barriers and facilitators to yoga among predominantly racial/ethnic minority, low-income adults. Complement Ther Clin Pract 2017;29:97–104.

21. Federation of State Medical Boards. Model guidelines for the use of complementary and alternative therapies in medical practice. 2002. Available from: http://www.fsmb.org/siteassets/advocacy/policies/model-guidelines-for-the-use-of-complementary-and-alternative-therapies-in-medical-practice.pdf. Accessed January 15, 2019.

22. Patel SJ, Kemper KJ, Kitzmiller JP. Physician perspectives on education, training, and implementation of complementary and alternative medicine. Adv Med Educ Pract 2017;8:499–503.

23. Burton A. Should your patient be doing yoga? Lancet Neurol 2014;13:241–242.

24. Qaseem A, Wilt TJ, McLean RM, Forciea MA. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2017;166:514–530.

25. Falsafi N. A randomized controlled trial of mindfulness versus yoga: effects on depression and/or anxiety in college students. J Am Psychiatr Nurses Assoc 2016;22:483–497.

26. Prathikanti S, Rivera R, Cochran A, Tungol JG, Fayazmanesh N, Weinmann E. Treating major depression with yoga: a prospective, randomized, controlled pilot trial. PLoS One 2017;12:e0173869.

27. Saeed SA, Antonacci DJ, Bloch RM. Exercise, yoga, and meditation for depressive and anxiety disorders. Am Fam Physician 2010;81:981–986.

28. Kwok JY, Choi KC, Chan HY. Effects of mind-body exercises on the physiological and psychosocial well-being of individuals with Parkinson's disease: A systematic review and meta-analysis. Complement Ther Med 2016;29:121–131.

29. Boulgarides LK, Barakatt E, Coleman-Salgado B. Measuring the effect of an eight-week adaptive yoga program on the physical and psychological status of individuals with Parkinson's disease. A pilot study. Int J Yoga Ther 2014;24:31–41.

30. Abrantes AM, Friedman JH, Brown RA, et al. Physical activity and neuropsychiatric symptoms of Parkinson disease. J Geriatr Psychiatry Neurol 2012;25:138–145.

31. Kwok JYY, Kwan JCY, Auyeung M, et al. Effects of mindfulness yoga vs stretching and resistance training exercises on anxiety and depression for people with Parkinson disease: A randomized clinical trial. JAMA Neurol 2019;76:755–763.

32. Babbar S, Hill JB, Williams KB, Pinon M, Chauhan SP, Maulik D. Acute feTal behavioral response to prenatal yoga: a single, blinded, randomized controlled trial (TRY yoga). Am J Obstet Gynecol 2016;214:399.e1–e8.

33. Babbar S, Shyken J. Yoga in pregnancy. Clin Obstet Gynecol 2016;59:600–612.

34. Polis RL, Gussman D, Kuo YH. Yoga in pregnancy: An examination of maternal and fetal responses to 26 yoga postures. Obstet Gynecol 2015;126:1237–1241.

35. Hartley L, Dyakova M, Holmes J, et al. Yoga for the primary prevention of cardiovascular disease. Cochrane Database Syst Rev 2014;(5):CD010072.

36. Pascoe MC, Thompson DR, Ski CF. Yoga, mindfulness-based stress reduction and stress-related physiological measures: a meta-analysis. Psychoneuroendocrinology 2017;86:152–168.

37. Wang W, Keh HT, Bolton LE. Lay theories of medicine and a healthy lifestyle. J Consum Res 2010;37:80–97.

38. Boehnke KF. Pain management: assembling a tool kit, building a life. JAMA 2018;320:2201–2202.

39. Ross A, Williams L, Pappas-Sandonas M, Touchton-Leonard K, Fogel D. Incorporating yoga therapy into primary care: the Casey Health Institute. Int J Yoga Ther 2015;25:43–49.

40. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care 2013;51:S30–S37.

41. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible sources of bias in primary care electronic health record data use and reuse. J Med Internet Res 2018;20:e185.

42. Hollister B, Bonham VL. Should electronic health record-derived social and behavioral data be used in precision medicine research? AMA J Ethics 2018;20:E873–E880.

43. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013;20:144–151.