Sleep Medicine 71 (2020) 66–76

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Original Article

Brief digital sleep questionnaire powered by machine learning prediction models identifies common sleep disorders

Alan R. Schwartz [a,b,c,*], Mairav Cohen-Zion [d,e], Luu V. Pham [a,b], Amit Gal [f], Mudiaga Sowho [a], Francis P. Sgambati [a,b], Tracy Klopfer [a], Michelle A. Guzman [a], Erin M. Hawks [a], Tamar Etzioni [g,h], Laura Glasner [e,i], Eran Druckman [j], Giora Pillar [g,h]

[a] Johns Hopkins Sleep Disorders Center, Baltimore, MD, USA
[b] Johns Hopkins Center for Interdisciplinary Sleep Research and Education, Baltimore, MD, USA[1]
[c] University of Pennsylvania Perelman School of Medicine, USA
[d] The Academic College of Tel Aviv-Jaffa, Tel Aviv, Israel
[e] DayZz Live Well Ltd, Herzeliya, Israel
[f] The Open University, Raanana, Israel
[g] Carmel Medical Center, Haifa, Israel
[h] Technion School of Medicine, Haifa, Israel
[i] Sheba Medical Center, Ramat Gan, Israel
[j] Druckman Research and Statistics, Rishon Lezion, Israel

ARTICLE INFO

Article history:
Received 9 October 2019
Received in revised form
17 February 2020
Accepted 5 March 2020
Available online 23 March 2020

Keywords:
Sleep disorders
Screening survey
Digital sleep questionnaire
Machine learning
Prediction model

ABSTRACT

Introduction: We developed and validated an abbreviated Digital Sleep Questionnaire (DSQ) to identify common societal sleep disturbances including insomnia, delayed sleep phase syndrome (DSPS), insufficient sleep syndrome (ISS), and risk for obstructive sleep apnea (OSA).
Methods: The DSQ was administered to 3799 community volunteers, of which 2113 were eligible and consented to the study. Of those, 247 were interviewed by expert sleep physicians, who diagnosed ≤2 sleep disorders. Machine Learning (ML) trained and validated separate models for each diagnosis. Regularized linear models generated 15–200 features to optimize diagnostic prediction. Models were trained with five-fold cross-validation (repeated five times), followed by robust validation testing. ElasticNet models were used to classify true positives and negatives; bootstrapping optimized probability thresholds to generate sensitivities, specificities, accuracies, and area under the receiver operating curve (AUC).
Results: Compared to reference subgroups, physician-diagnosed sleep disorders were marked by DSQ evidence of sleeplessness (insomnia, DSPS, OSA), sleep debt (DSPS, ISS), airway obstruction during sleep (OSA), blunted circadian variability in alertness (DSPS), sleepiness (DSPS and ISS), increased alertness (insomnia) and global impairment in sleep-related quality of life (all sleep disorders). ElasticNet models validated each diagnosis with high sensitivity (80–83%), acceptable specificity (63–69%), high AUC (0.80–0.85) and good accuracy (agreement with physician diagnoses, 68–73%).
Discussion: A brief DSQ readily engaged and efficiently screened a large population for common sleep disorders. Powered by ML, the DSQ can accurately classify sleep disturbances, demonstrating the potential for improving the sleep, health, productivity and safety of populations.

1. Introduction

It is estimated that nearly all adults experience transient sleep disturbances over the course of their lives, and that these disturbances persist and remain chronic in up to 40% of the population at large [1]. Several factors contribute to the development and maintenance of sleep disturbances including environmental,

* Corresponding author. Johns Hopkins Sleep Disorders Center, Baltimore, MD, USA
E-mail address: aschwar02@gmail.com (A.R. Schwartz).
[1] www.cisre.jhu.edu

https://doi.org/10.1016/j.sleep.2020.03.005
1389-9457/

biological, medical and psychological factors, which can predispose, precipitate or perpetuate a cycle of worsening sleep and daytime function [2,3]. Common sleep disorders including insomnia, sleep apnea, circadian rhythm disturbances and chronic insufficient sleep constitute a major health and societal burden, contributing to a loss of personal well-being, the development of chronic disease, lost work productivity and performance, and spikes in accidents that result in needless injury and death [4–15]. Efforts to stem the tide of chronic sleep disorders across society, however, have languished from a lack of efficient means for accessing diagnostic acumen within the medical community.

Gaps in the societal awareness and medical training can impede the recognition of common sleep disorders in society at large. Sleep competes with a full agenda of personal and occupational goals, activities and behaviors that de-emphasize its importance and requirement for satisfying a fundamental biologic need [16,17]. In this context, the exigencies of daily life conspire to make sleep seem expendable, and may interfere with behaviors conducive to optimizing sleep patterns. Medical and psychiatric conditions also contribute to sleep disturbances, which can in turn accelerate the progression of these underlying disorders. Societal awareness of sleep disorders is further compromised by a lack of medical experts in sleep medicine who are trained in a specialty that is highly interdisciplinary rather than insular. Primary care providers are hampered by a lack of basic knowledge or training in sleep medicine and are overextended with other health maintenance priorities that constrain and compete with their time and attention [1,18]. Moreover, the health consequences and societal impact of sleep disorders has not been fully appreciated [17,19], which impedes recognition in the medical setting and lay public.

In recent years, the demands of living in an increasingly global 24/7 society can conflict with natural sleep and circadian rhythms, contributing to widespread sleep complaints and frank sleep disorders [20,21]. In response, the public has sought help in maintaining predictable, healthy, regular sleep patterns over time. The industry has attempted to address this need with the proliferation of novel activity trackers that offer sleep insights and peeks consumers' interest in monitoring and optimizing their sleep patterns. The recent surge in proliferation of wearable devices has spurred consumers to monitor and characterize their sleep disturbances without formally engaging medical professionals. A full complement of validated paper-and-pencil survey instruments have also been developed to assess for specific sleep disorders including sleep apnea [22], insomnia [23], circadian rhythm disturbances [24,25] and insufficient sleep [26–30]. Nonetheless, a simple, easy to use, yet comprehensive survey instrument is still required to identify common sleep disorders in those who complain of sleep disturbances in the public at large.

The present study was designed to address a critical unmet need for well-validated, accurate, scalable and time-efficient approaches that allow people to identify and characterize their sleep disorders without having to seek medical attention. We hypothesized that a simple screening sleep survey presented via a digital platform could: (1) readily engage a large segment of the adult population with undiagnosed sleep disorders, and (2) predict gold-standard sleep diagnoses generated from expert physicians trained in sleep medicine. To test these hypotheses, we developed and tested a brief online survey instrument that probed for several common sleep disorders in community volunteers, and validated this approach for predicting specific sleep disorders against gold-standard diagnoses from experts in sleep medicine. To achieve this goal, we applied modern machine learning (ML) methods to train and test the performance of four separate statistical models, one for each diagnosis of interest: insomnia, delayed sleep phase syndrome (DSPS), insufficient sleep syndrome (ISS), and suspected obstructive sleep apnea (OSA). Our findings suggest that a short, online survey instrument can classify common sleep disorders in people from the general community who report sleep disturbances, and could swiftly assign tentative diagnoses and greatly expedite clinical care.

2. Methods

2.1. Participants

Participants with sleep concerns aged 20–65 years were recruited for our protocol. Advertisements were directed through Facebook to users in the Baltimore metropolitan area, who responded with specific self-reported sleep complaints or normal sleep (Fig. 1). A total of 3799 people responded to advertisements, of which 805 were ineligible based on age ($n = 85$), use of sleep aids on more than three times per week ($n = 166$), pregnancy or breast feeding ($n = 41$), night shift work ($n = 143$), or a previously diagnosed sleep disorder or polysomnographic sleep study ($n = 370$). Another 880 persons chose not to consent for the protocol, leaving a total of 2114 consenting eligible participants who completed the Digital Survey Questionnaire (DSQ, see Supplement). Of the remaining subjects, a convenience sample of 252 subjects were invited for an in-person interview, of which five did not show up for their appointment. The final sample included 247 participants with a spectrum of sleep-related symptoms who agreed to a face-to-face interview at the Johns Hopkins Sleep Center. The entire protocol was approved by the Johns Hopkins Medical Center Institutional Review Board, and informed consent was obtained from each participant.

2.2. Study protocol

Subjects who were invited to the sleep center completed a structured two-step face to face interview process. Participants were interviewed by a trained research associate who inquired systematically about sleep/wake patterns, nocturnal sleep disruption and sleep apnea symptoms, medications and chronic medical conditions. Subjects' sleep, medical and social history was then reviewed by highly experienced sleep medicine physicians (AS, LVP, MS). The physician then interviewed the participant and recorded up to two sleep-related diagnoses (Fig. 1). A nominal fee was provided for study subjects to compensate them for their time and travel expenses. All study data were captured electronically in a HIPAA secured REDCap database that was customized and maintained by the Johns Hopkins Center for Interdisciplinary Sleep Research and Education (www.cisre.jhu.edu).

2.3. Sleep survey instruments (digital sleep questionnaire, DSQ)

A condensed questionnaire was designed to survey a broad range of sleep patterns and symptoms. Investigators (GP, TE) initially compiled a series of questions that targeted several commonly encountered sleep disturbances in the general community including insomnia, delayed sleep phase syndrome, insufficient sleep and obstructive sleep apnea. Survey questions were selected based on clinical intuition and were refined through ad hoc pilot inquiry from patients with sleep disorders. The final survey for the present study encompassed questions focusing on several sleep/wake domains including:

- **Insomnia** when participants reported difficulties initiating or maintaining sleep, possibly accompanied by anxiety and/or depression.

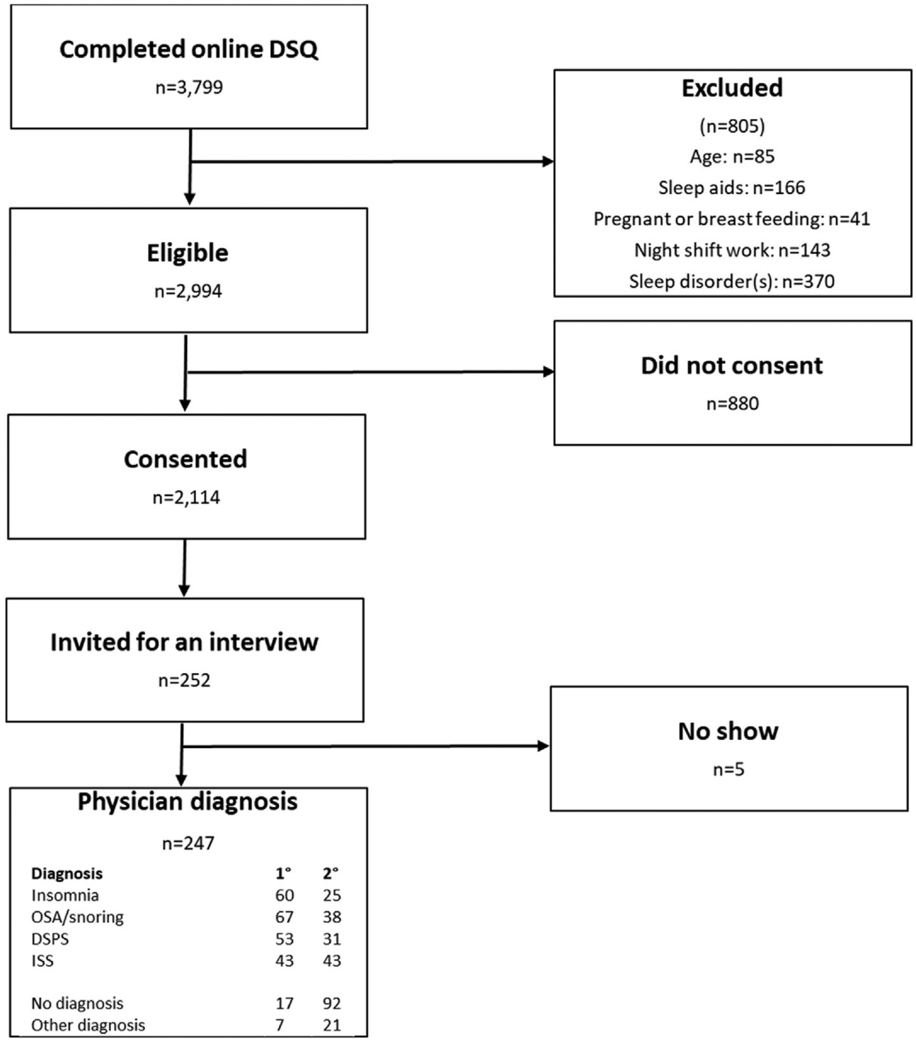

Fig. 1. Consort diagram illustrating on-line recruitment yield among eligible consenting adults in the general community, a portion of whom were interviewed face-to-face by sleep physicians. Primary (1°) and secondary (2°) sleep diagnoses are represented in those who underwent face-to-face interviews.

- **Delayed sleep phase** when subjects reported a delayed bedtime on weekends compared to weekdays, and/or differences in alertness in the evening compared to morning hours.
- **Insufficient sleep syndrome** when participants reported excessive sleep debt, which was defined as the difference in sleep duration between weekend and weekdays, as well as the difference between perceived sleep need and actual sleep duration.
- **Obstructive sleep apnea** when subjects reported loud, frequent snoring and/or witnessed apneic episodes at night, often combined with excessive daytime sleepiness.

As with most questionnaire-based approaches, concern exists that self-report data could be biased, inaccurate or prone to several threats of validity. Our methodologic approach addresses these concerns in three ways. First, we used domain experts to configure and validate the content of the questionnaire. Second, we conducted analyses for internal validity with the commonly used Cronbach alpha statistic to check for correlations among the variables. Assessing for randomness or high variance exists across subjects with this metric, we found high values (ie, higher levels of correlation among variables and subjects), suggesting consistency in subjects' responses to the DSQ and statistical evidence reliability. Cronbach's alpha for the variables included in our models were 0.76, 0.78, 0.70 and 0.77 for sleep apnea, insomnia, delayed sleep phase and insufficient sleep, respectively fell within an acceptable range in the literature. Third, self-reported responses were checked for their predictive validity, (ie, whether they predicted objective conclusions [diagnoses] of sleep medicine experts). Our findings indicate that the DSQ rendered accurate predictions and could discriminate common sleep disorders, thus providing further evidence for the validity of this questionnaire.

The entire Digital Sleep Questionnaire (DSQ) appears in Supplement. In addition, subjects completed the Epworth Sleepiness Scale (ESS) and Functional Outcomes of Sleep Questionnaire (FOSQ).

2.4. Statistical analysis

Analyses were structured to address the primary hypothesis that DSQ responses predicted the sleep diagnoses generated by expert sleep physicians. Two approaches were taken to address this hypothesis. First, DSQ responses were described among subgroups, based on sleep experts' primary diagnoses. ANOVA and Chi-squared analysis were used to discern between-group differences in continuous outcomes and distributions of responses (proportions) in categorical responses, respectively. To determine the source of differences in continuous variables, a Bonferroni correction was implemented for post-hoc testing involving multiple comparisons. Statistical significance was inferred from a $p < 0.05$.

Second, a machine learning (ML) approach was employed to develop, optimize and validate predictive models (Fig. 2). This process consisted of three steps: (1) feature engineering, (2) model fitting and (3) threshold setting, as detailed below. Each step included a validation phase to verify that outcomes were likely to be generalizable, and free from common threats to validity such as overfitting and inadequate power.

2.4.1. Feature Engineering

Participants' DSQ responses provided raw data. In conventional analysis, one typically aims to assess how a measured construct (eg, an item or group of items in the questionnaire that represent a meaningful response) is related to an outcome. In predictive modeling, however, our aim is to predict the outcome without concern for the individual effects of each response. In practice, raw data were "sliced and diced" into "features", each representing a specific type of information from which new features could be further built from basic ones. For example, a response on a 5-point Likert scale could be split into several features (eg, response = 4, ≥2, <3, or even no response to handle missing values). Then two or more such basic features could be combined to create features like "at least one of feature A and feature B are true" with features A and B assembled from previously generated features.

In the current study, we used all Likert scale responses (18 in total) as a basis for feature engineering. For each such response (A, B, C,...), all features that were generated took the form of: response = X, response < X and response > X. We then generated all the features, which took the form: "both A and B", "A or B" and "A or B, but not both". This systematic process yielded more than 20,000 features containing detailed information from participants' responses. Due to the myriad of possible combinations of features generated by this approach, this process did not encompass all possible features from the DSQ dataset. However, our objective was to build sufficiently accurate predictive models from the features we generated from the current approach, which fulfilled this expectation. These features also remained understandable, logical, and interpretable since they could be easily reconstructed from participants' original responses, making the resulting prediction models that much more transparent and interpretable.

When generating such large numbers of features, concern arose about a loss of statistical power from a surfeit of feature and multicollinearity among feature, both of which could result in models that are overfit or lack generalizability. This approach also risked generating features with little bearing on desired outcomes. To reduce these risks, we performed three tests on each feature and discarded features that failed to meet criteria for any of these tests:

(1) Test of variance: The features we generated were binary ("dummy variables"). To pass the variance test, the feature must have values of zero for at least 5% of our sample size, and values of 1 at least 5% of our sample. Implementing these minimum thresholds ensured that the features we created were not restricted to specific cases and had sufficient predictive potential.
(2) Test of covariance: If two features were <5% different from each other, we discarded one of them.
(3) Test for "Lift": Lift is a measure of how well a binary variable helps predict a target variable. If the tested variable was X and the target was Y, lift was the ratio between the expectation of Y under the condition of X = 1, and the expectation of Y under the condition of X = 0. Unless stated otherwise, the formula for lift was: $LIFT(x,y) = E(Y|x = 1)/E(Y|x = 0)$. Since this measure used the target variable, statistical artifacts could occur from any specific subsample of our data. To minimize the risk of a spurious result, we generated 200 random samples from our data (without replacement, sample sizes were 50% of the original sample size) and computed

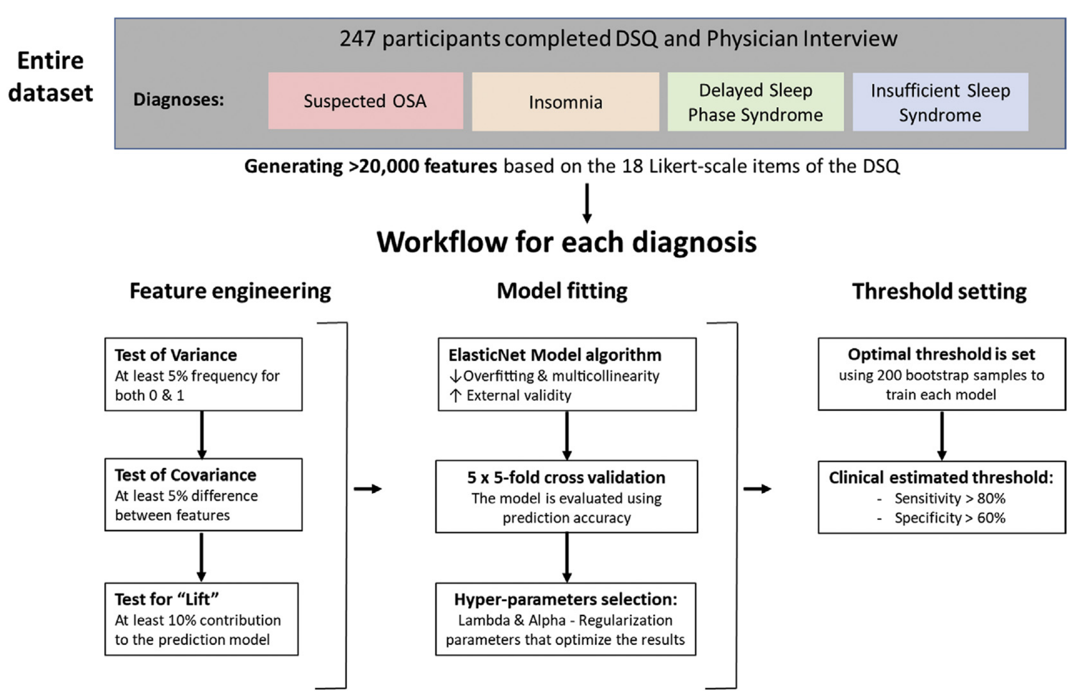

Fig. 2. Overview of machine learning approach that was applied to predicting physician-diagnosed sleep disorders. The approach encompassed three major phases of model development and validation as shown and described in greater detail in text.

200 point estimates of lift from the random samples to estimate the distribution of lift values. We first checked that the overall lift of the feature lay within 1 standard deviation of the mean for this distribution. We then computed a 95% confidence interval for this distribution and selected only features whose confidence interval was above 1.1 (ie, 10% increase) or below 0.9 (ie, 10% decrease). In other words, we chose features only if lift was likely to improve our prediction by at least 10%.

Lift testing depended on chosen outcome (ie, physician diagnosis). Since we sought to predict each of the four diagnoses, we built four datasets of features, one for each prediction model. Each model contained between ~300 and ~1100 features (with some features appearing in several models).

2.4.2. Model fitting

To build a predictive model, we used an "ElasticNet" approach. ElasticNet is very similar to logistic regression in that the model assigns a coefficient to every feature and then uses these coefficients to compute a score representing the log odds of the target diagnosis. The training data was used to estimate the best set of coefficients. These coefficients were then used to assess the log odds (or probabilities) for new data (test data) as well. ElasticNet models, however, incorporate algorithms that differ from conventional logistic regression when computing coefficients, making it better suited to predict rather than explain the potential impact of each predictor on the diagnosis (in contrast to standard regression, which focuses on estimating the magnitude of associations between predictor and outcome variables; see further detail in Discussion).

Specifically, ElasticNet algorithms incorporated two "regularization mechanisms", whose purpose was to reduce the risks of overfitting and multicollinearity, thereby increasing generalizability of the models. These mechanisms included the "Lasso" (or L_1 regularization) and "Ridge" (or L_2 regularization) methods, as previously described [31]. In short, the Lasso mechanism aggressively prunes out features that do not contribute to predictive power by penalizing coefficients with large magnitude, while the ridge mechanism shrinks the coefficients of highly correlated features, thus reducing the variance of the predictions and the impact of potential multicollinearity. These techniques, widely used in building ElasticNet models, were deployed to minimize the above-mentioned risks and thereby enhance our predictive accuracy [32].

Prediction accuracy with ElasticNet models was accomplished with a 5-fold cross validation approach. In short, we split our data into 5 groups, and repeated the model fitting process 5 times. Each time, 4 groups were extracted to "train" the model, and the predictive accuracy was then measured on the fifth group to which the model remained blinded during training. In every iteration, the model was fit to a different training set, and hence the resulting coefficients varied from one iteration to another. Furthermore, our target measure, which focused on the model's performance in the testing set was different in each iteration. Rather than validating a specific set of coefficients, our approach validated the process of obtaining the coefficients. Repeating the process multiple times (each time with different training and testing datasets) yielded the average test error, which provided a reliable estimate of the full model's performance on future datasets that were encountered. This cross-validation process ensured the external validity of the final model.

Our ElasticNet models depended on two "hyper-parameters" that governed how strongly the two regularization mechanisms influenced the resulting model [33]. One parameter, lambda (ranging from 0 to infinity), controlled to what extent the overall regularization of predictors altered the coefficients in the model (with 0 representing no regularization at all, and infinity representing the highest level of regularization in which all the coefficients in the model are set to zero). Another parameter, alpha, controlled the ratio between the two above-mentioned regularization mechanisms. We optimized the choice of these two parameters through the cross-validation process. That is, we repeated the cross-validation process many times with different values of lambda and alpha and compared the cross-validation average error before selecting those parameters yielding the greatest accuracy for the model.

2.4.3. Threshold setting

There ElasticNet models, like logistic regression models, produced probability estimates for the desired diagnostic outcome. Although diagnostic tools generally express probabilities for a given diagnosis on a scale of 0–1, we chose to report more commonly used measures of sensitivity and specificity for the purpose of medical diagnosis. For this reason, we set a threshold that allowed us to convert intermediate probabilities into yes/no diagnoses. Specifically, if the predicted probability exceeded a given threshold, the predictive model returned a positive diagnosis (ie, predicted a 1 rather than a 0); otherwise, the model predicted the absence of this diagnosis (ie, predicted a 0 rather than a 1). Hard fast diagnostic predictions enabled us to compute each model's specificity and sensitivity, and the area under the receiver operating curve (AUC).

In order to choose an optimal threshold, we used 200 bootstrap samples, trained our model on each sample, and generated probability predictions on cases that were left out. For each candidate threshold, we used these predictions to compute sensitivity and specificity, and estimated the mean (±SD) for each threshold. For purposes of diagnostic prediction, we sought thresholds with an estimated sensitivity >80% and specificity >60%, if possible.

As mentioned, interpretation of the actual coefficients for each feature was limited by the fact that the features themselves combined elements from the original data, and the ElasticNet algorithm shrunk the coefficients in order to increase predictive power. Nonetheless, we generated an estimate of the importance of each feature and reported the composition of these features by diagnosis (see Supplement), based on their components from original DSQ variables. The importance estimates were normalized so that 100% represented the most important feature, and all other levels of importance were set relative to this feature. These estimates provided a semi-quantitative guide to the impact of specific features and their underlying components on the model's predictive accuracy. Accuracy was defined as the percent agreement between the physician and model prediction for each diagnosis.

3. Results

3.1. Subject characteristics

Participants' characteristics are summarized for the entire group by physicians' primary sleep diagnoses in Table 1.

Women predominated across all diagnoses, and most especially with chronic insufficient sleep, consistent with epidemiologic surveys on sleep patterns [34]. The majority were in stable relationships with a partner, largely employed, non-Hispanic Caucasian, middle-aged individuals. Subjects with insomnia, sleep phase delay and chronic insufficient sleep were generally overweight to mildly obese (based on BMI). Those with suspected sleep apnea, however, were moderate to severely obese and had elevated neck, waist and hip circumferences compared to the other reference groups.

Table 1
Demographic and anthropometric characteristics, overall and by diagnosis.

Characteristic	OSA/Snoring (n = 67)	Insomnia (n = 60)	DSPS (n = 53)	ISS (n = 43)	None (n = 17)	Other (n = 7)	Total (N = 247)
Gender							
Female	36 (53.7%)	36 (60%)	33 (62.3%)	30 (69.8%)	10 (58.8%)	4 (57.1%)	149 (60.3%)
Male	31 (46.3%)	24 (40%)	20 (37.7%)	13 (30.2%)	7 (41.2%)	3 (42.9%)	98 (39.7%)
Marital status							
Single	16 (23.9%)	15 (25%)	17 (32.1%)	9 (20.9%)	3 (17.6%)	1 (14.3%)	61 (24.7%)
Married/Relationship	48 (71.6%)	37 (61.6%)	27 (50.9%)	29 (67.4%)	14 (82.4%)	6 (85.7%)	161 (65.2%)
Separated	0 (0%)	2 (3.3%)	3 (5.7%)	2 (4.7%)	0 (0%)	0 (0%)	7 (2.8%)
Divorced	3 (4.5%)	5 (8.3%)	5 (9.4%)	3 (7.0%)	0 (0%)	0 (0%)	16 (6.5%)
Widowed	0 (0%)	1 (1.7%)	1 (1.9%)	0 (0%)	0 (0%)	0 (0%)	2 (8.0%)
Employment							
Unemployed	8 (11.9%)	5 (8.3%)	5 (9.4%)	2 (4.7%)	0 (0%)	0 (0%)	20 (8.1%)
Employed	59 (88.1%)	55 (91.7%)	48 (90.6%)	41 (95.3%)	17 (100%)	7 (100%)	227 (91.9%)
Ethnicity							
Non-Hispanic	62 (92.5%)	56 (93.3%)	48 (90.6%)	37 (86.0%)	16 (94.1%)	7 (100%)	226 (91.5%)
Hispanic	5 (7.5%)	4 (6.7%)	5 (9.4%)	6 (14.0%)	1 (5.9%)	0 (0%)	21 (8.5%)
Race							
Caucasian	55 (82.1%)	45 (75%)	37 (69.8%)	35 (81.4%)	15 (88.2%)	5 (71.4%)	192 (77.7%)
African American	5 (7.5%)	9 (15%)	4 (7.5%)	1 (2.3%)	1 (5.9%)	1 (14.3%)	21 (8.5%)
Asian	4 (6.0%)	4 (6.7%)	10 (18.9%)	5 (11.6%)	1 (5.9%)	0 (0%)	24 (9.7%)
Other	3 (4.5%)	2 (3.3%)	2 (3.8%)	2 (4.7%)	0 (0%)	1 (14.3%)	10 (4.0%)
Age (years)	44.4 (11.0)	44.2 (12.4)	35.8 (11.8)	35.0 (11.8)	33.2 (10.9)	38.6 (9.5)	39.9 (12.4)
BMI (kg/m^2)	34.6 (7.7)	29.6 (6.9)	26.6 (5.9)	27.1 (44)	25.2 (3.4)	32.9 (10.9)	29.7 (7.3)
Neck circumference (cm)	38.6 (4.6)	35.7 (4.2)	34.7 (3.8)	34.4 (3.5)	34.0 (3.8)	37.0 (2.9)	36.0 (4.4)[a]
Waist circumference (cm)	106.6 (19.0)	94.9 (17.6)	85.3 (13.3)	86.6 (12.7)	85.2 (11.3)	111.0 (40.0)	94.4 (19.2)[a]
Hip circumference (cm)	117.2 (17.4)	108.8 (14.1)	103.5 (11.6)	105.0 (9.8)	102.0 (10.5)	114.4 (19.0)	109.0 (14.9)[a]

OSA: obstructive sleep apnea; Snoring: habitual snoring; DSPS: delayed sleep phase syndrome; ISS: insufficient sleep syndrome. Other diagnoses included hypersomnia (n = 4) and unclassified (n = 3).
[a] missing: 2 cases.

3.2. Sleep characteristics

Sleep characteristics reported by participants on the DSQ are summarized in Table 2 and were largely consistent with the clinical features of the underlying sleep disturbances as follows.

Participants showed evidence of significant sleep debt among all diagnostic categories ($p = 0.001$). Sleep debt was approximately 45 min greater in subgroups with primary diagnoses of delayed sleep phase and insufficient sleep syndromes than in those with insomnia ($p = 0.058$ and 0.017, respectively) or presumed sleep apnea ($p = 0.008$ and 0.017, respectively). As expected, increasing severity of insomnia was endorsed by those who received this primary diagnosis from expert physicians as compared to those without or with other sleep disorders ($p < 0.0001$). Mild to moderate insomnia was also reported in those with a sleep phase delay and suspected sleep apnea. Sleep apnea was most commonly suspected in those who snored loudly ($p < 0.0001$) and frequently ($p < 0001$), and was least suspected in those not reporting having witnessed apneic episodes ($p < 0.001$) as compared to subgroups without or with other sleep diagnoses. Circadian variability in daytime alertness was observed in those diagnosed with insufficient sleep, insomnia and suspected sleep apnea. In contrast, those diagnosed with a sleep phase delay expressed no decided diurnal difference in alertness levels between morning and evening hours ($p < 0.001$). The Functional Outcomes of Sleep Questionnaire (FOSQ), a global measure of sleep-related quality of life, was significantly higher in those without compared to all those with a sleep diagnosis ($p < 0.0001$). Significant differences were detected in Epworth Sleepiness Scale (ESS) score among diagnostic groups ($p < 0.0001$) with elevations noted in those with sleep phase delay (4.6 ± 0.9, $p < 0.0001$) and insufficient sleep (7.1 ± 1.0, $p < 0.0001$) compared to those with no primary sleep diagnosis. Of note, the ESS was 3.7 ± 1.0 lower in the subgroup with a primary diagnosis of insomnia than that with insufficient sleep ($p = 0.003$), even though both groups reported comparable sleep duration on weekdays.

3.3. Predictive models by clinical diagnosis

Features are listed for each of the following models by their relative importance in predicting the physicians' diagnoses (see Supplement). For each sleep disorder, model performance was represented by predictiveness plots, which examined the relationship between the average prevalence of the specific disorder in our cohort as function of the model's estimated probability for obtaining this outcome. These graphs demonstrate a monotonic rise in prevalence as the estimated probability for each disorder rose, indicating that diagnoses “made” by these ElasticNet models correlated well with the expert physician-generated diagnoses (see Supplement).

In characterizing the diagnostic performance of each model, we found the sensitivity, specificity, accuracy and AUC to be in the good to excellent range, as summarized in Table 3 and in **Results** immediately below. Values for sensitivity and specificity were used to define Receiver Operating Curves for each of the four prediction models (Fig. 3). Each graph contains two curves. The cross-validation ROC curve was based on the models created during the cross-validation process. Because this process resampled the data multiple times, each point on the graph reflected multiple predictions, yielding a smooth curve. The final models consisted of only one prediction value per observation, as represented by the prediction ROC curve. This curve was used to generate the final AUC and accuracy reported in Table 3. Because this curve was based on fewer predictions, the resulting curve was a step line.

Table 2
Selected DSQ and sleep-related parameters for primary sleep diagnoses.

	OSA/Snoring (n = 67)	Insomnia (n = 60)	DSPS (n = 53)	ISS (n = 43)	None (n = 17)	Other (n = 7)	Total (n = 247)
Amount of sleep needed (min)							
	472.4 (82.2)	455.8 (73.7)	471.2 (63.4)	465.4 (77.8)	468.5 (46.7)	480.0 (98.0)	466.8 (73.6)
Weekday sleep duration (min)							
	406.8 (91.9)	356.8 (72.1)	389.2 (77.3)	370.1 (62.3)	441.2 (41.8)	366.4 (89.9)	385.7 (79.6)
Weekend sleep duration (min)							
	462.8 (77.6)	413.8 (80.6)	486.5 (75.1)	478.3 (100.7)	493.2 (56.1)	505.7 (110.1)	462.0 (86.4)
Sleep debt* (min)							
	56.0 (75.1)	57.0 (61.5)	97.4 (86.5)	108.1 (97.2)	52.1 (45.7)	139.3 (108.4)	76.3 (81.6)
Insomnia severity (past month)							
Mild	7 (10.4%)	13 (21.7%)	10 (18.9%)	5 (11.6%)	1 (5.9%)	2 (28.6%)	38 (15.4%)
Moderate	16 (23.9%)	22 (36.7%)	14 (26.4%)	3 (7%)	0 (0%)	1 (14.3%)	56 (22.7%)
Severe	5 (7.5%)	10 (16.7%)	2 (3.8%)	0 (0%)	0 (0%)	0 (0%)	17 (6.9%)
Very severe	1 (1.5%)	3 (5%)	2 (3.8%)	0 (0%)	0 (0%)	0 (0%)	6 (2.4%)
No insomnia	38 (56.7%)	12 (20.0%)	25 (47.2%)	35 (81.4%)	16 (94.1%)	4 (57.1%)	130 (52.6%)
Self-reported nocturnal breathing pauses							
Never	6 (9%)	13 (21.7%)	17 (32.1%)	17 (39.5%)	13 (76.5%)	0 (0%)	66 (26.7%)
Rarely	3 (4.5%)	8 (13.3%)	8 (15.1%)	2 (4.7%)	2 (11.8%)	1 (14.3%)	24 (9.7%)
Occasionally	13 (19.4%)	2 (3.3%)	5 (9.4%)	4 (9.3%)	1 (5.9%)	0 (0%)	25 (10.1%)
Often	5 (7.5%)	0 (0%)	1 (1.9%)	0 (0%)	0 (0%)	0 (0%)	6 (2.4%)
Very often	6 (9%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (2.4%)
I don't know	34 (50.7%)	37 (61.7%)	22 (41.5%)	20 (46.5%)	1 (5.9%)	6 (85.7%)	120 (28.6%)
Snoring							
Never	1 (1.5%)	6 (10%)	11 (20.8%)	7 (16.3%)	11 (64.7%)	0 (0%)	36 (14.6%)
Rarely	4 (6%)	18 (30%)	12 (22.6%)	9 (20.9%)	5 (29.4%)	3 (42.9%)	51 (20.6%)
Occasionally	12 (17.9%)	21 (35%)	17 (32.1%)	16 (37.2%)	0 (0%)	1 (14.3%)	67 (27.1%)
Often	14 (20.9)	6 (10%)	5 (9.4%)	5 (11.6%)	0 (0%)	1 (14.3%)	31 (12.6%)
Very often	33 (49.3%)	3 (5%)	4 (7.5%)	3 (7%)	0 (0%)	1 (14.3%)	44 (17.8%)
I don't know	3 (4.5%)	6 (10%)	4 (7.5%)	3 (7%)	1 (5.9%)	1 (14.3%)	18 (7.3%)
Loud snoring							
No	9 (13.4%)	34 (56.7%)	36 (67.9%)	29 (67.4%)	16 (94.1%)	3 (42.9%)	127 (51.4%)
Yes	58 (86.6%)	26 (43.3%)	17 (32.1%)	14 (32.6%)	1 (5.9%)	4 (57.1%)	120 (48.6%)
Circadian preference							
Morning	14 (20.9%)	20 (33.3%)	9 (17.0%)	18 (41.9%)	8 (47.1%)	1 (14.3%)	70 (28.3%)
Evening	23 (34.3%)	19 (31.7%)	7 (13.2%)	12 (27.9%)	4 (23.5%)	2 (28.6%)	67 (27.1%)
Neither	30 (44.8%)	21 (35%)	37 (69.8%)	13 (30.2%)	5 (29.4%)	4 (57.1%)	110 (44.5%)
FOSQ score*	91.6 (16.8)	96.5 (16.3)	94.5 (15.4)	91.6 (15.6)	110.5 (11.8)	92.4 (23.6)	94.8 (16.6)
ESS score	8.7 (4.3)	6.4 (4.0)	7.8 (4.4)	10.3 (5.3)	3.8 (3.7)	11.1 (7.3)	7.9 (4.8)

*Difference between weekday and weekend sleep duration; *Missing 3 cases; Self-reported breathing events: witnessed apneic episodes, nocturnal snorting, gasping or choking; OSA: obstructive sleep apnea; Snoring: habitual snoring; DSPS: delayed sleep phase syndrome; ISS: insufficient sleep syndrome; FOSQ: Functional Outcomes of Sleep Questionnaire; ESS: Epworth Sleepiness Scale.

3.3.1. *Sleep Apnea Prediction Model*

Our model for the diagnosis of suspected sleep apnea started with 1110 features, of which 289 were chosen by the ElasticNet algorithm. Snoring (DSQ #4) was the most important factor, contributing to 198 fine-grained predictive features. In addition, ease of waking up on weekday mornings (DSQ #18) contributed to 23 features, followed by early morning awakenings (DSQ #16) and likely to fall asleep in a passive/comfortable state (DSQ #26) along with additional predictive features of lesser importance (see Supplement). The overall sensitivity of the model was 83.4 ± 7.4% (±SEM) and specificity was 66.5 ± 7.6%, based on a threshold of 0.3, and the AUC for the receiver operating curve was 0.85. The overall accuracy of our model was 73.2 ± 4.4%.

Table 3
Performance parameters of prediction models by diagnosis.

Parameter	OSA	Insomnia	DSPS	ISS
Sensitivity	83.4%	80.3%	80.5%	82.3%
Specificity	66.5%	69.4%	62.9%	63.6%
Accuracy	73.2%	72.9%	67.9%	69.6%
AUC[a]	0.85	0.83	0.80	0.82

[a] Area under Curve (Receiver Operating Curve).

3.3.2. *Insomnia model*

Our model for the diagnosis of Insomnia started with 1119 features, of which 280 were chosen by the ElasticNet algorithm to predict this diagnosis. Our model incorporated three factors, which contribute similarly to predicting a diagnosis of insomnia. These factors included early morning awakenings (DSQ #16), nocturnal awakenings (DSQ #32) and snoring. (DSQ #4), followed by factors of lower importance including the likelihood of dozing in a passive/uncomfortable state (DSQ #23), ease in waking up on weekday mornings (DSQ #18) and feeling anxious (DSQ #25), as detailed in Supplement. The overall sensitivity of this prediction model was 80.3 ± 8.0% and the specificity was 69.4 ± 6.5%, based on a threshold of 0.3. The AUC for the ROC was 0.83, yielding an overall accuracy of 72.9 ± 4.1%.

3.3.3. *Delayed Sleep Phase Syndrome Model*

This model started with 1123 features, of which only 21 where chosen by the ElasticNet algorithm for the predictive model. These features incorporated some sleep/wake patterns and psychological components, which captured a lack of circadian fluctuations in alertness (morningness/eveningness, DSQ #24), difficulty in performing complex tasks (DSQ #22), depressed mood (DSQ #27) and a delayed weekend bedtime (DSQ #3). See Supplement for details. The overall sensitivity was 80.4 ± 9% and the specificity was 62.9 ± 8.6%, which were achieved with a threshold of 0.25 and

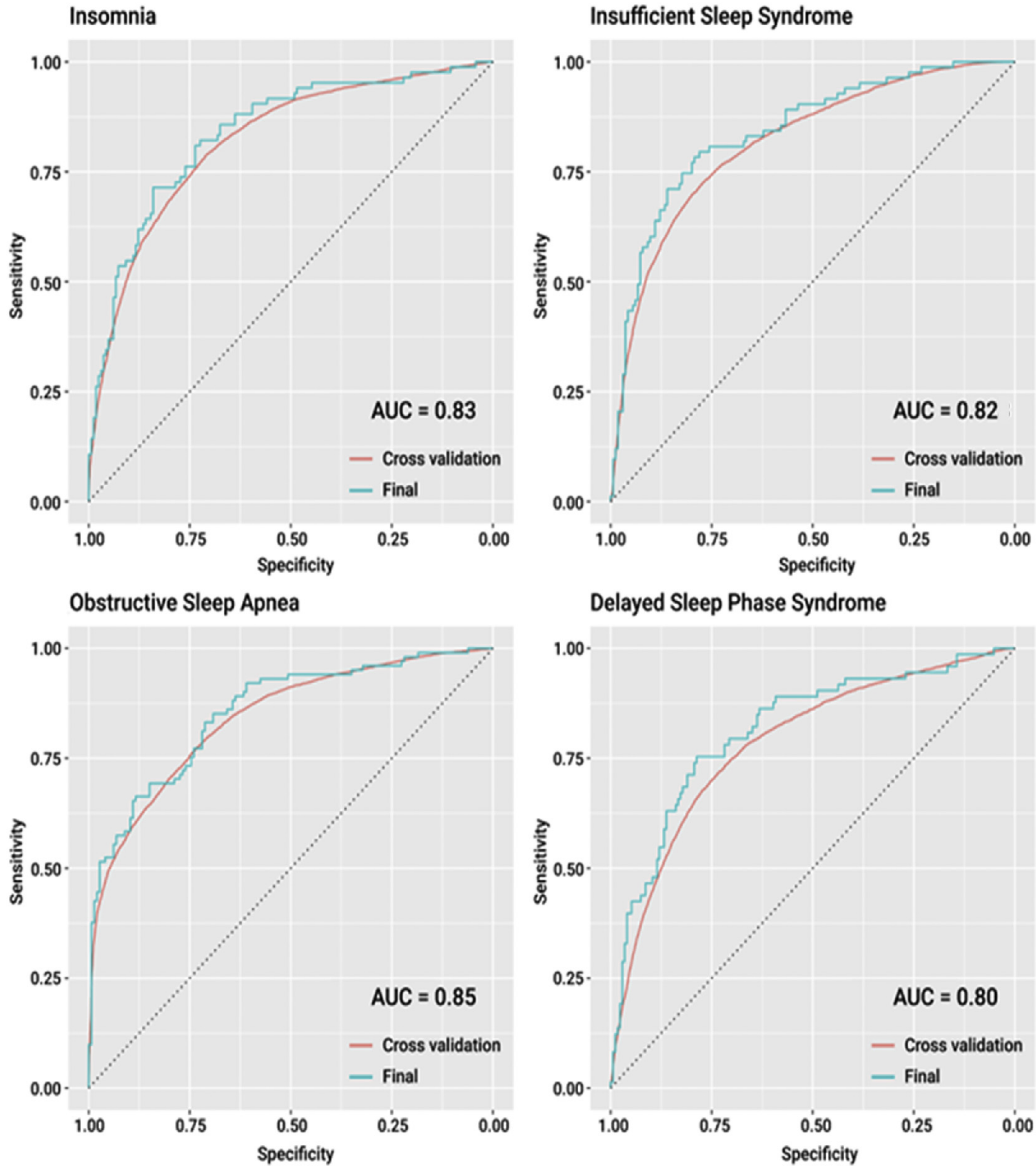

Fig. 3. Receiver Operating Curves (ROC) are illustrated for each of four prediction models, based on data generated from the cross-validation process and the final model. See text for details.

yielded an AUC for the ROC curve of 0.80. The overall accuracy of our model was 67.9 ± 5%.

3.3.4. *Insufficient Sleep Syndrome Model*

For this model, 294 features were initially identified, of which 96 were incorporated by the ElasticNet algorithm. Predictive factors encompassed important features that included a high likelihood of dozing in a passive state or uncomfortable position (DSQ #23), early morning awakenings (DSQ #16), nocturnal awakenings (DSQ #31), feeling unrefreshed in the morning (DSQ #17) along with difficulty awakening on a weekday morning (DSQ #18). See Supplement for details. The overall sensitivity of this predictive model was 82.2 ± 7.5% and the specificity was 63.6 ± 7.5%, based on a threshold of 0.25, which yielded an AUC for the ROC of 0.82. The overall accuracy of our model was 69.6 ± 4.7%.

4. Discussion

Our findings demonstrated that a simple, abbreviated digital sleep questionnaire (DSQ) can be readily deployed in those who suffer sleep disturbances across the general adult community. It can identify common sleep disorders with a high degree of accuracy that closely approximates diagnoses rendered by expert sleep physicians. Leveraging state of the art ML statistical methods, we achieved high sensitivity with acceptable specificity in classifying sleep disorders in a community sample of individuals with insufficient sleep, circadian rhythm disturbances, insomnia and suspected obstructive sleep apnea. Of note, social media provided an efficient platform for canvassing several thousands of people, who willingly provided confidential demographic, anthropometric and sleep history online (under informed consent) without face to face interaction. Although these participants who reported sleep-related complaints, the vast majority had never sought medical attention to address these issues. The findings suggested that sleep disturbances constitute a large unrecognized unmet need across a broad segment of modern society, and that the DSQ offers means for accurately detecting and classifying underlying sleep disorders as a first step in targeting individuals for evaluation and management.

Sleep survey instruments have generally targeted specific sleep disorders in patients presenting to health care providers, who might or might not be trained in sleep medicine [22]. For example, several questionnaires have been developed to determine the probability that patients will test positive for obstructive sleep apnea including the Berlin Questionnaire [22], the Multivariable Apnea Prediction tool [35], and STOP-BANG survey [36,37], effectively triaging patients for overnight testing and characterizing peri-operative risk. These questionnaires generally target risk factors and symptoms of obstructive sleep apnea. In contrast, the DSQ integrates apnea detection with other common sleep disorders, effectively accelerating the screening for sleep disturbances in an unreferred, community-based population. Its availability online promises to empower those with sleep disturbances and their primary care providers with a streamlined approach for identifying the possibility of sleep apnea across large segments of the work force and general population.

The DSQ is unique in that it targets several sleep disturbances that are highly prevalent in society at large [38]. Short sleep is a common complaint that results from diverse factors including insufficient sleep, circadian rhythm misalignment and underlying psychosocial, physiologic and medical causes of insomnia. The DSQ distinguished specific categories of sleeplessness by emulating a clinician's intuition in eliciting a sleep history. For example, a diagnosis of insufficient sleep could be inferred from estimates of sleep requirements, actual sleep time and accumulated sleep debt [39,40]. Similarly, circadian tendencies were compared to the subject's actual sleep times to determine whether circadian misalignment could contribute to the subject's sleeplessness. Insufficient sleep and circadian rhythm disorders could both lead to chronic decreases in total sleep time, as distinguished from other types of insomnia in which sleeplessness involved lesser degrees of volitional control over sleep/wake patterns [41,42]. The DSQ probed these sleep/wake constructs to characterize several sleep disturbances in unreferred subjects all at once, instead of restricting its focus to a single sleep disorder or grading the severity of a single sleep disorder among sleep center referrals. The DSQ also offered an advantage over another abbreviated, broad-based well-recognized Global Sleep Assessment Questionnaire (GSAQ) in its focus on common causes of inadequate sleep in the general community (insufficient sleep, circadian rhythm disturbances, insomnia and sleep apnea) rather than primary sleep disorders that mainly present to primary care and sleep referral practices (such as periodic limb movements, restless legs syndrome and parasomnias). These distinct advantages serve to enhance the DSQ's applicability to people who suffer from sleep disturbances in the general community.

The utility of the DSQ was further enhanced by machine learning methods that were used to generate well-validated models for predicting common sleep disturbances. This approach was chosen over conventional logistic regression because it prioritized the goal of generating reliable diagnostic predictions rather than quantifying associations between specific predictors and diagnoses of interest. Our approach incorporated separate analyses of training and validation datasets, thereby embedding procedures that ensure the external validity of our models. Data from the study sample were split so that models could be fit repeatedly to training datasets and applied thereafter to predict outcomes in the validation datasets. ML helped us circumvent other pitfalls of conventional regression analysis in that regularization of predictors reduced the risk of overfitting and multi-collinearity in multidimensional datasets with attendant reductions in statistical power (as described in Methods). A further advantage of our approach is that ML generated “features” (predictor variables), which often took the form of non-linear combinations of underlying raw data elements from the DSQ (as described in Supplement). Procedures for optimizing the selection of these features balanced improvements in the models' performances in predicting the diagnoses of interest (as assessed by reductions in the overall models' variances), albeit sacrificing some increase in bias [31]. Implementing a ML approach, we generated highly sensitive models for predicting common sleep disturbances with moderate specificity from an abbreviated, easily deployed survey of subjects' characteristics and sleep/wake patterns. Similarly, high levels of accuracy can be anticipated when these models are applied in people with sleep disturbances from the general population. Nonetheless, we acknowledge that the DSQ was developed from a largely symptomatic segment of the population, and its performance cannot necessarily be used to screen the entire population or those presenting to a clinical setting. The DSQ's generalizability as a screening instrument would require further prospective testing in larger, more varied populations of interest.

Several limitations should be considered in interpreting our findings. First, our survey focused primarily on sleep disturbances that are commonly found in the general population and did not encompass less prevalent sleep disorders that can be found in referral populations, such as narcolepsy, idiopathic hypersomnia, movement disorders and parasomnias. Second, while our sample was small, it was derived from a community sample of people who reflected a large pool of persons reporting sleep symptoms and disturbances. We acknowledge that our relatively small sample size increased the risk of overfitting our ML prediction models. Although regularization methods were implemented to control for this risk and our approach yielded good results, larger datasets might still be required to confirm and further “train” the model, thereby increasing their accuracy, external validity and generalizability. Third, predicting sleep diagnoses with ElasticNet models can limit mechanistic insight to be gained from these models because (1) computed coefficients for model features can be difficult to interpret, and the magnitude of these coefficients is often reduced by regularization of predictor variables, which are expressed as proportions of the variables' distributions around the means rather than as absolute magnitudes. Mechanistic insight is also limited by the fact that (2) ElasticNet models do not yield p values, precluding customary statistical inferences about associations between specific features (predictors) and outcomes (diagnoses). Fourth, our predictive models were constructed from a convenience sample rather than a systematic population sample (from which the actual prevalence of each sleep disorder could be derived in the general community). Nonetheless, this pragmatic approach yielded solid prediction models that can be applied to self-identified individuals with sleep concerns. Fifth, we modeled both the physician-assigned primary and secondary sleep diagnoses since people often present with combinations of sleep disorders that can be interrelated (eg, sleep phase delay or circadian misalignment, both of which can result in chronic insufficient sleep). Sixth, although we modeled the potential for sleep apnea in our subjects, we did not confirm this diagnosis with objective nocturnal testing. Yet, our approach was designed to identify subjects considered worthy of sleep testing by sleep medicine physicians. Seventh, the DSQ was designed for application to the general population of adults, who could have had concomitant medical, pharmacologic and psychosocial co-morbidities. The DSQ was highly sensitive in detecting diagnoses rendered by our sleep experts, making it an appropriate screening tool for those who suffer from sleep disturbances in the community at large. Its specificity, although comparable to those reported for clinical sleep survey instruments (eg, STOP-BANG, GSAQ), would suggest that its use is most appropriate in classifying sleep disorders in those with sleep-related symptoms, and should not be extended to broad segments of the population indiscriminately.

Our study has major implications for the future practice of sleep medicine. It offers evidence that sleep disturbances constitute a large unmet need in the general population, long before subjects are inclined to address their concerns in a medical setting. We implemented a unique approach for predicting common sleep disorders, based on experts' clinical intuition in the survey's development, and based on ML methods for analyzing DSQ responses. To date, most sleep symptom surveys utilize straightforward scoring methods to summarize the presence and severity of specific sleep disorders. In contrast, our approach to analyzing DSQ responses complements clinical intuition by generating novel features that markedly improved the survey's predictive ability. In so doing, the DSQ offers a readily deployable, scalable online instrument that can empower the lay public to take control of its own sleep health and respond to sleep disturbances with minimal effort and no additional sensors. In combination with wearable technologies, the DSQ can identify people at high risk for sleep disorders, thereby targeting appropriate management strategies. Our study offers a novel approach for extending the clinician's reach with insights that have the potential for improving the sleep, health, productivity and safety of the general populations [43–46]. We nonetheless emphasize that the utility and deployment of the DSQ in the general population will require careful consideration of its performance characteristics and its purpose and impact in specific settings and work milieus. Depending on the setting in which the DSQ is deployed, its specificity can be traded to maximize sensitivity when detection of sleep disorders impacts health, quality of life, productivity and public safety. The overall approach heralds an abundance of opportunities for supervised machine learning and sleep research that combines clinical and artificial intelligence for the betterment of patients and society at large.

Summary

An abbreviated digital sleep questionnaire (DSQ) powered by machine learning can accurately detect and classify common sleep disturbances in the general community as a means for developing public health strategies to identify and mitigate sleep disorders.

Impact

The DSQ offers a readily deployable, scalable and accurate means for screening large populations with the potential for improving the sleep, health, productivity and safety of entire populations.

CRediT authorship contribution statement

Alan R. Schwartz: Conceptualization, Project administration, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Mairav Cohen-Zion:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Luu V. Pham:** Methodology, Investigation, Writing - review & editing. **Amit Gal:** Software, Validation, Formal analysis, Visualization, Writing - review & editing. **Mudiaga Sowho:** Investigation. **Francis P. Sgambati:** Resources, Software, Data curation. **Tracy Klopfer:** Investigation, Supervision. **Michelle A. Guzman:** Investigation, Supervision. **Erin M. Hawks:** Investigation, Supervision, Project administration. **Tamar Etzioni:** Conceptualization, Investigation. **Laura Glasner:** Visualization, Writing - review & editing. **Eran Druckman:** Formal analysis. **Giora Pillar:** Conceptualization, Methodology, Investigation.

Acknowledgements

This work was supported by dayZz Live Well Ltd (Herzliya, Israel).

Conflict of interest

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: https://doi.org/10.1016/j.sleep.2020.03.005.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sleep.2020.03.005.

References

[1] Altevogt BM, Colten HR, editors. Sleep disorders and sleep deprivation: an unmet public health problem; 2006. Retrieved from, www.nap.edu/catalog/11617.html.

[2] Smolensky MH, Di Milia L, Ohayon MM, et al. Sleep disorders, medical conditions, and road accident risk. Accid Anal Prev 2011;43(2):533–48. https://doi.org/10.1016/J.AAP.2009.12.004.

[3] Wong ML, Lau EYY, Wan JHY, et al. The interplay between sleep and mood in predicting academic functioning, physical health and psychological health: a longitudinal study. J Psychosom Res 2013;74(4):271–7. https://doi.org/10.1016/j.jpsychores.2012.08.014.

[4] Albarrak M, Banno K, Sabbagh AA, et al. Utilization of healthcare Resources in obstructive sleep apnea syndrome: a 5-year follow-up study in men using CPAP. Sleep 2005;28(10):1306–11. Retrieved from, https://watermark.silverchair.com/sleep-28-10-1306.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAm0wggJpBgkqhkiG9w0BBwagggJaMIICVgIBADCCAk8GCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMQdafhD4AflkHOSrBAgEQgIICIL9L8dM_KIvFnHpDIHe9zP8eLrcSNR6j8cpTmH.

[5] Bjornsdottir E, Keenan BT, Eysteinsdottir B, et al. Quality of life among untreated sleep apnea patients compared with the general population and changes after treatment with positive airway pressure. J Sleep Res 2015;24(3):328–38. https://doi.org/10.1111/jsr.12262.

[6] Cappuccio FP, D'Elia L, Strazzullo P, et al. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2010;33(2):414–20. https://doi.org/10.2337/dc09-1124.

[7] Carskadon MA. Sleep deprivation: health consequences and societal impact. Med Clin 2004;88(3):767–76. https://doi.org/10.1016/j.mcna.2004.03.001.

[8] Chien KL, Chen PC, Hsu HC, et al. Habitual sleep duration and insomnia and the risk of cardiovascular events and all-cause death: report from a community-based cohort. Sleep 2010;33(2):177–84. https://doi.org/10.1093/sleep/33.2.177.

[9] Garbarino S, Guglielmi O, Sanna A, et al. Risk of occupational accidents in workers with obstructive sleep apnea: systematic review and meta-analysis. Sleep 2016;39(6):1211–8. https://doi.org/10.5665/sleep.5834.

[10] Gottlieb DJ, Ellenbogen JM, Bianchi MT, et al. Sleep deficiency and motor vehicle crash risk in the general population: a prospective cohort study. BMC Med 2018;16(1):44. https://doi.org/10.1186/s12916-018-1025-7.

[11] Knauert M, Naik S, Gillespie MB, et al. Clinical consequences and economic costs of untreated obstructive sleep apnea syndrome. World J Otorhinolaryngol Head Neck Surg 2015;1(1):17–27. https://doi.org/10.1016/J.WJORL.2015.08.001.

[12] Liu Y, Wheaton AG, Chapman DP, et al. Sleep duration and chronic diseases among US adults age 45 years and older: evidence from the 2010 behavioral risk factor surveillance system. Sleep 2013;36(10):1421–7. https://doi.org/10.5665/sleep.3028.

[13] Rosekind MR, Gregory KB, Mallis MM, et al. The cost of poor sleep: workplace productivity loss and associated costs. J Occup Environ Med 2010;52(1):91–8. https://doi.org/10.1097/JOM.0b013e3181c78c30.

[14] Swanson LM, Arnedt JT, Rosekind MR, et al. Sleep disorders and work performance: findings from the 2008 National Sleep Foundation Sleep in America poll. J Sleep Res 2011;20(3):487–94. https://doi.org/10.1111/j.1365-2869.2010.00890.x.

[15] Wang Y, Mei H, Jiang YR, et al. Relationship between duration of sleep and hypertension in adults: a meta-analysis. J Clin Slp Med 2015;11(9):1047–56. https://doi.org/10.5664/JCSM.5024.

[16] Barnes CM, Wagner DT, Ghumman S. Borrowing from sleep to pay work and family: expanding time-based conflict to the broader nonwork domain. Person Psychol 2012;65(4):789–819. https://doi.org/10.1111/peps.12002.

[17] Luyster FS, Strollo PJ, Zee PC, et al. Sleep: a health imperative. Sleep 2012;35(6):727–34. https://doi.org/10.5665/sleep.1846.

[18] Miller CM. Lack of training in sleep and sleep disorders. AMA J Ethics 2008;10(9):560–3. https://doi.org/10.1001/virtualmentor.2008.10.9.medu1-0809.

[19] Grandner MA, Malhotra A. Sleep as a vital sign: why medical practitioners need to routinely ask their patients about sleep. Sleep Health 2015;1(1):11–2. https://doi.org/10.1016/j.sleh.2014.12.011.

[20] Groeger JA, Zijlstra FRH, Dijk DJ. Sleep quantity, sleep difficulties and their perceived consequences in a representative sample of some 2000 British adults. J Sleep Res 2004;13(4):359–71. https://doi.org/10.1111/j.1365-2869.2004.00418.x.

[21] Heath AC, Eaves LJ, Kirk KM, et al. Effects of lifestyle, personality, symptoms of anxiety and depression, and genetic predisposition on subjective sleep disturbance and sleep pattern. Twin Res Hum Genet 1998;1(4):176–88. Retrieved from, http://www.stockton-press.co.uk/tr.

[22] Netzer NC, Stoohs RA, Netzer CM, et al. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med 1999;131(7):485–91. https://doi.org/10.7326/0003-4819-131-7-199910050-00002.

[23] Morin CM, Belleville G, Bélanger L, et al. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011;34(5):601–8. https://doi.org/10.1093/sleep/34.5.601.

[24] Horne JA, Ostberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol 1976;4(2):97–110. Retrieved from, http://www.ncbi.nlm.nih.gov/pubmed/1027738.

[25] Roenneberg T, Kuehnle T, Juda M, et al. Epidemiology of the human circadian clock. Sleep Med Rev 2007;11(6):429–38. https://doi.org/10.1016/J.SMRV.2007.07.005.

[26] Berrios GE, Shapiro CM. ABC of sleep disorders." I don't get enough sleep, doctor". BMJ Br Med J 1993;306(6881):843–6. https://doi.org/10.1136/bmj.306.6881.843.

[27] Centers for Disease Control and Prevention (CDC). Percentage of adults who reported an average of ≤6 hours of sleep per 24-hour period, by sex and age group—United States 1985 and 2004. MMWR. Morbidity and mortality weekly report, vol. 54. Centers for Disease Control and Prevention (CDC); 2005.

[28] Centers for Disease Control and Prevention (CDC). Unhealthy sleep-related behaviors–12 States, 2009. MMWR. Morbidity and mortality weekly report, vol. 60; 2011. Retrieved from, http://www.ncbi.nlm.nih.gov/pubmed/21368738.

[29] Liu Y, Njai RS, Greenlund KJ, et al. Relationships between housing and food insecurity, frequent mental distress, and insufficient sleep among adults in 12 US states, 2009. Prev Chronic Dis 2014;11:E37. https://doi.org/10.5888/pcd11.130334.

[30] Roth T, Zammit G, Kushida C, et al. A new questionnaire to detect sleep disorders. Sleep Med 2002;3(2):99–108. https://doi.org/10.1016/S1389-9457(01)00131-9.

[31] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning data: mining, inference, and prediction. 2nd ed. 2009. https://doi.org/10.1007/978-0-387-84858-7.

[32] Kim M-H, Banerjee S, Park SM, et al. Improving risk prediction for depression via elastic net regression - results from korea national health insurance services data. In: AMIA annual symposium proceedings. vol. 2016; 2016. p. 1860–9. Retrieved from, http://www.ncbi.nlm.nih.gov/pubmed/28269945.

[33] Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019;19(1):64. https://doi.org/10.1186/s12874-019-0681-4.

[34] Lauderdale DS, Knutson KL, Yan LL, et al. Objectively measured sleep characteristics among early-middle-aged adults. Am J Epidemiol 2006;164(1):5–16. https://doi.org/10.1093/aje/kwj199.

[35] Maislin G, Pack AI, Kribbs NB, et al. A survey screen for prediction of apnea. Sleep, vol. 18; 1995. Retrieved from, https://academic.oup.com/sleep/article-abstract/18/3/158/2749647.

[36] Chung F, Abdullah HR, Liao P. STOP-bang questionnaire: a practical approach to screen for obstructive sleep apnea. Chest 2016;149(3):631–8. https://doi.org/10.1378/chest.15-0903.

[37] Proczko MA, Stepaniak PS, De Quelerij M, et al. STOP-Bang and the effect on patient outcome and length of hospital stay when patients are not using continuous positive airway pressure. J Anesth 2014;28(6):891–7. https://doi.org/10.1007/s00540-014-1848-0.

[38] Klingman KJ, Jungquist CR, Perlis ML. Questionnaires that screen for multiple sleep disorders. Sleep Med Rev 2017;32:37–44. https://doi.org/10.1016/j.smrv.2016.02.004.

[39] Chattu V, Manzar M, Kumary S, et al. The global problem of insufficient sleep and its serious public health implications. Healthcare 2019;7(1):1. https://doi.org/10.3390/healthcare7010001.

[40] Van Dongen HPA, Maislin G, Mullington JM, et al. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 2003;26(2):117–26. https://doi.org/10.1093/sleep/26.2.117.

[41] Chesson A, Hartse K, Anderson WM, et al. Practice parameters for the evaluation of chronic insomnia. An American academy of sleep medicine report. Standards of practice committee of the American academy of sleep medicine. Sleep 2000;23(2):237–41. Retrieved from, http://www.ncbi.nlm.nih.gov/pubmed/10737341.

[42] Roth T, Roehrs T. Insomnia: epidemiology, characteristics, and consequences. Clin Cornerstone 2003;5(3):5–15. Retrieved from, http://www.ncbi.nlm.nih.gov/pubmed/14626537.

[43] Centers for Disease Control and Prevention (CDC). Short sleep duration among workers–United States, 2010. MMWR. Morbidity and mortality weekly report, vol. 61; 2012. Retrieved from, http://www.ncbi.nlm.nih.gov/pubmed/22534760.

[44] Hillman D, Mitchell S, Streatfeild J, et al. The economic cost of inadequate sleep. Sleep 2018;41(8). https://doi.org/10.1093/sleep/zsy083.

[45] Hillman DR, Murphy AS, Pezzullo L. The economic cost of sleep disorders. Sleep 2006;29(3):299–305. https://doi.org/10.1093/sleep/29.3.299.

[46] Webb WB. The cost of sleep-related accidents: a reanalysis. Sleep 1995;18(4):276–80. https://doi.org/10.1093/sleep/18.4.276.