Using Explainable Deep Learning and Logistic Regression to Evaluate Complementary and Integrative Health Treatments in Patients with Musculoskeletal Disorders

Douglas F. Redd, PhD
Washington DC VAMC
George Washington University
doug_redd@gwu.edu

Joseph L. Goulet, PhD, MS
VA Connecticut Healthcare System
Yale University
Joseph.Goulet@va.gov

Yijun Shao, PhD
VA Salt Lake City Health Care System
George Washington University
yshao@gwu.edu

Cynthia A. Brandt, MD, MPH
VA Connecticut Healthcare System
Yale University
Cynthia.Brandt@va.gov

Qing Zeng-Treitler, PhD
Washington DC VAMC
George Washington University
zengq@gwu.edu

Abstract

There is an increasing interest in the use of Complementary and Integrative Health (CIH) for treatment of pain as an alternative to opioid medications. We use a novel explainable deep learning approach compared and contrasted to a traditional logistic regression model to explore the impact of musculoskeletal disorder related factors on the use of CIH. The impact scores from the neural network show high correlation with the log-odds ratios of the logistic regression, showing the promise that neural networks can be used to identify high impact factors without depending on a priori assumptions and limitations of traditional statistical models.

1. Introduction

There is increasing interest in the use of Complementary and Integrative Health (CIH) approaches for pain as an alternative to opioid medications. Opioid misuse, abuse, addiction, and overdose are a worldwide public health crisis [1]. In the US, opioid prescription-related deaths have quadrupled since 1999, and are a leading cause of death today [2]. While men are more likely to die from opioid overdose than women, that gap is closing [2]. Patients with mental health and substance use disorders, in particular those with post-traumatic stress disorder (PTSD), have higher rates of opioid-related morbidity and mortality [3, 4]. Currently, there is intense interest in alternatives to opioids for treating pain [5-7]. However, there is an urgent need to identify interventions that can reduce opioid initiation, without increasing harms, and to identify factors associated with their utilization.

The Institute of Medicine (IOM) noted that a large and growing array of evidence-based CIH including acupuncture, massage, meditation, and yoga, among others, may hold special appeal to persons with pain [8]. CIH has been used for pain and a number of other conditions [9]. Results from a survey of Veterans with chronic non-cancer pain (N=401) indicated that 82% reported some prior CIH use, and 99% a willingness to try it [10]. According to another recent study, Operation Enduring Freedom (OEF)/Operation Iraqi Freedom (OIF)/Operation New Dawn (OND) Veterans, women, and younger Veterans are more likely to use CIH [11].

Although studies directly comparing CIH and opioids are lacking [12], recent systematic reviews reveal that when compared to a common control group of usual care, the CIH modalities to be studied here have a similar magnitude of pain improvement [13-18]. However, the evidence base regarding the impact of CIH on opioid use is limited. “No study to date has evaluated the effectiveness of select CIH interventions for reducing opioid initiation, or for reducing opioid or stopping use.” [19]

Our team has been conducting a larger project that aims to study the effect of CIH use on pain management including opioid use and whether the effect varies by demographic and/or clinical characteristics. As part of the project, we set out to identify characteristics associated with the utilization of CIH in patients with chronic pain, employing both deep learning and logistic

URI: https://hdl.handle.net/10125/64140
978-0-9981331-3-3

regression approaches. Logistic regression is commonly used in clinical and health services research, though it makes some assumptions: a linear relationship between the logit of the outcome and predictor variables, no extreme values or outliers in the continuous predictors and no high intercorrelations among the predictor variables. Deep learning, like many machine learning algorithms, does not make such assumptions and is capable of modeling non-linear relationships. Furthermore, we have been developing an explainable artificial intelligence (AI) method called Impact Assessment (IA) that allows the use of deep learning to study factors, including patient demographic and clinical characteristics as well as facility attributes, that are associated with the use of CIH.

We have used Veteran Affairs (VA) Electronic Health Record (EHR) data in the analysis. VA is the largest integrated healthcare system in the US and a leader in quality and efficiency due in part to the transformation of its EHR [20-22]. Most EHR data are in readily analyzable structured data fields (e.g. weight, stop codes, ICD, and CPT codes) [21]. However, many services do not have CPT or ICD codes, and CIH receipt may be detailed in clinical notes [23]. Structured data alone are insufficient to identify CIH use [24]. For example, patients may obtain CIH from non-VA providers, particularly when a specific therapy is not available at a local VA facility. However, VA providers often document them in clinical notes. In this work, we developed a natural language processing tool to extract CIH use from clinical notes with area under the curve (AUC) ranging from 82.8% to 91.8%, depending on the CIH modalities.

2. Methods

2.1 Dataset and CIH Extraction

We used data from the VA Musculoskeletal Disorders (MSD) Cohort, developed by Goulet et al [25]. The cohort contains data on patients with back, neck, and large joint disorder diagnoses. The key domains and sample variables in the MSD cohort are shown in Table 1.

To extract the CIH utilization documented in clinical notes of members of the MSD cohort, we annotated a random sample of clinical notes from the VA EHR stored in the Veterans Administration Informatics and Computing Infrastructure (VINCI) database. A team of clinical experts defined a list of CIH-related keywords for different CIH modalities, which was used to identify CIH-related notes. For example, for acupuncture the keywords were "acupuncture", "ACUP", and "needling". From CIH-related notes, snippets composed of the keywords together with 30 words before and 30 words after were extracted. We chose 30 words in response to experience in previous studies on the same EHR data set [26, 27, 28, 29]. The snippets were categorized into 6 modalities: Acupuncture, Biofeedback, Guided Imagery [30], Meditation, Tai-Chi and Yoga. For each modality, a small subset (n=500~600) of the snippets was selected for human annotation. The human annotated data were then used to develop, train and test NLP classifiers.

Table 1. Key domains and example variables contained in the MSD cohort.

Domain	Variables
Demographics	DOB, race/ethnicity, sex, service-connected status, service era
Diagnoses	Specific MSD(s), PTSD, substance use disorders, etc.
Procedures	CPT and ICD, including for CIH
Vitals	Pain intensity NRS scores, BP, height/weight
Consults & referrals	Specialty pain clinic, opioid substitution therapy (e.g. methadone)
Pharmacy	Opioids, tramadol, gabapentin, antidepressants, benzodiazepines
Health screening	Smoking, alcohol, PTSD, and depression screen results
Laboratory data	Urine drug test, liver function test, and pathology results
Risk factors	Prior overdose or suicide-related event
Treatment	Mental health clinic stop codes

An annotation guideline was developed and iteratively refined through group chart review. The annotation labels were "current user," "planned/recommended," "uncertain," "past user," and "none-user." We further grouped the original multiple category annotation labels into binary labels: "current user" (positive) vs. "all other cases" (negative). "Current user" means that the snippet shows a patient was a current CIH user at the time when the note was written. The annotation was first performed by a dedicated annotator and subsequently reviewed according to the guideline and revised by 2 other team members. Questions and disagreements were resolved by consensus.

We first tokenized the snippets by converting all upper-case characters to lower case and removing all punctuation and numbers. We then generated both 1-gram and 2-gram bag of words features. The 1-gram

features were unique words and 2-gram features were two adjacent words that were originally (i.e., before removing punctuations and numbers) separated by only white spaces. Because of the large number of 2-gram features, we selected 2-grams based on their discriminative power.

We trained support vector machine (SVM) models to classify the snippets. After experimenting with several other kernels such as a polynomial kernel, we selected a linear kernel as it yielded the best performance. We used 10-fold cross validation to measure the classification performance, i.e. splitting the annotated data into 10 subsets, using 9 subsets for training and 1 subset for testing, and repeating the process 10 times.

We calculated performance metrics including area under the receiver operating characteristics curve (AUC) and accuracy, with AUC being the primary metric and accuracy a secondary metric. The SVM classifiers were optimized for AUC in all experiments. The final performances were micro-averaged over the 10-folds for each metric. The AUC ranged from 82.8% to 91.8% (Table 2).

In all, 26,769,725 document snippets having CIH keywords were identified from 17,072,822 distinct documents, belonging to 15,095,504 visits, in turn representing 2,283,936 individual patients. We randomly sampled 10,000 patients to study the demographic and clinical factors associated with CIH use. Characteristics of the patients are presented in Table 3.

We selected 19 variables in 5 categories as predictors (Table 4). These represented demographics (age and gender), race/ethnicity, vitals (pain, body mass index (BMI)), comorbidities (post-traumatic stress disorder (PTSD), major depressive disorder (MDD), bipolar disorder, anxiety disorder, hypertension, coronary artery disease (CAD), and Charlson comorbidity index (CCI)), behaviors (drug use disorder, alcohol use disorder, and smoking history), and analgesic use (opioid and non-opioid). The variables were selected based on their high prevalence among Veterans in VHA care, and their known or hypothesized association with CIH.

Data was aggregated by patient, with age calculated as the average of each patient's age at the time of the visits corresponding to the sampled documents. Analgesic use was determined by counting the number of filled opioid or non-opioid analgesic prescriptions for a patient, where the prescription fill occurred within one month prior to or following one of the CIH correlated visits. 239 patients were excluded due to missing values for multiple variables, resulting in 9,761 patients in the analysis. In addition, 851 patients were missing BMI values. These missing BMI values were assigned the average of the BMI of the other patients in the analysis. The final 9,761 patients randomly assigned to sets for training (60%), validation (20%), and testing (20%).

2.2 Deep Neural Network

In order to identify characteristics associated with the utilization of CIH, and to assess an alternative to logistic regression, we chose to implement a Deep Neural Network (DNN). However, a concern with DNN methods is that they are viewed as black-boxes and are

Table 2. Performance metrics for CIH modality classification.

Modality	AUC	Accuracy
Acupuncture	0.918	0.858
Biofeedback	0.864	0.804
Guided Imagery	0.910	0.854
Meditation	0.828	0.833
Tai-Chi	0.878	0.818
Yoga	0.856	0.798

Table 3. Demographics and clinical characteristics of the patient sample

All Patients	**9,761**
Basic Demographics	
Male	8,805 (90%)
Female	956 (10%)
Age: Mean, Median	57, 58
Race/Ethnicity	
White	6,472 (66%)
Black	2,114 (22%)
Hispanic	615 (6%)
Other/Unknown	560 (6%)
Vitals	
Pain: Mean, Median	3.9, 4
BMI: Mean, Median	29.2, 28.5
Comorbidities	
PTSD	1,704 (17%)
MDD	1,012 (10%)
Bipolar Disorder	544 (6%)
Anxiety Disorder	1,134 (12%)
Hypertension	4,152 (43%)
CAD	1,182 (12%)
CCI: Mean, Median	0.68, 0
Drug Use Disorder	981 (10%)
Alcohol Use Disorder	1,533 (16%)
Smoking History	6,598 (68%)
Analgesic Use	
Opioid	3,781 (39%)
Non-Opioid	3,704 (38%)
CIH Use	5,001 (51%)

difficult to interpret, yet interpretability is critical for clinical adoption of findings. One approach to explain AI models is Local Interpretable Model-Agnostic Explanations (LIME) [31]. The Impact Assessment method we use is related to the LIME approach. We have validated the Impact Assessment method by comparing its results with odds ratios derived from logistic regression, finding strong but not perfect correlations as expected [32]. Another validation approach is to use a dataset with known underlying relationship between predictors and outcome. Since we do not have the complete knowledge of underlying relationships in real patient datasets, we experimented with simulated datasets with 100 variables and a mixture of linear and non-linear relationships. Noting that simulated data are far less complex that real patient datasets, we observed that the Impact Assessment is highly accurate (90-95%) in estimating the contribution of a variable to the outcome.

In order to evaluate our Impact Assessment method using real, non-simulated data, we built a DNN using the CIH data from section 2.1. The DNN was constructed with an input layer of 19 nodes, corresponding to the 19 variables, and an output layer of a single node using a sigmoid activation function giving the probability of CIH use. Five fully connected hidden layers were used, all using the rectified linear unit function for activation. The number of nodes in the hidden layers were 200, 300, 200, 300, and 200. Weights were initialized with random small numbers and updated using stochastic gradient descent with Nesterov momentum. A mini-batch size of 100 was used, along with a learning rate of 0.001 and a momentum of 0.9. AUC was measured against the validation set after each epoch. We used 150 epochs, at which point no improvement in AUC was seen in the last 10 epochs. The final DNN model was kept for subsequent impact score measurements.

Table 4. CIH outcome and predictor variables

Outcome variable	Variable Type	Representation
CIH use	Dichotomous	0 = No CIH use, 1 = at least one CIH use
Predictors		
Demographics		
Age	Continuous	Normalized to 0 – 1
Gender	Dichotomous	0 = Male, 1 = Female
Race/Ethnicity		
White	Dichotomous	0 = Non-White, 1 = White
Black	Dichotomous	0 = Non-Black, 1 = Black
Hispanic	Dichotomous	0 = Non-Hispanic, 1 = Hispanic
Vitals		
Pain	Ordinal	Normalized to 0 – 1
BMI	Continuous	Normalized to 0 – 1
Comorbidities		
PTSD	Dichotomous	0 = no PTSD, 1 = PTSD
MDD	Dichotomous	0 = no MDD, 1 = MDD
Bipolar Disorder	Dichotomous	0 = no Bipolar, 1 = Bipolar
Anxiety Disorder	Dichotomous	0 = no Anxiety, 1 = Anxiety
Hypertension	Dichotomous	0 = no Hypertension, 1 = Hypertension
CAD	Dichotomous	0 = no CAD, 1 = CAD
CCI	Ordinal	Normalized to 0 – 1
Drug Use Disorder	Dichotomous	0 = no non-alcohol, non-tobacco drug abuse, 1 = non-alcohol, non-tobacco drug abuse
Alcohol Use Disorder	Dichotomous	0 = no Alcohol use, 1 = Alcohol use
Smoking History	Dichotomous	0 = no history of smoking, 1 = smoking history
Analgesic Use		
Opioid	Dichotomous	0 = no opioid use, 1 = opioid use
Non-opioid	Dichotomous	0 = no non-opioid analgesic use, 1 = non-opioid analgesic use

2.3 Impact Assessment

We define a new variable called impact score. For each variable, a reference value is selected. For binary variables such as diagnoses, 0 is viewed as the reference, as it has been used to indicate the absence of a diagnosis. For categorical variables (e.g. race), we select a category (e.g. unknown) based on convention and clinical context as the reference. For continuous variables, we use the mean as the reference. To calculate the impact score for a variable, we change its current value to the reference value and observe change in the outcome. If the outcome does not change, this suggests that the change in value has no impact. If the model changes, the impact score will be calculated as follows:

The last layer of the DNN outputs a value p between 0 and 1 through a sigmoid function $p = \sigma(x)$. The change of prediction originates from change in x. One way to obtain x from p is the logit function:

$$x = \text{logit}(p) = \log\frac{p}{1-p}$$

Therefore, the individual-level impact score is defined as:

$$\frac{\text{logit}(p_{\text{cur}}) - \text{logit}(p_{\text{ref}})}{(\text{current value}) - (\text{reference value})}$$

where: p_{ref} is the new value of p after changing the current value to reference value. Note that the score is only defined if the current value differs from the reference value.

We define the impact score at the population level simply as the mean of all impact scores of patients on whom the score is defined. The Impact score can be interpreted similarly to the log-odds ratio from logistic regression. It represents the average rate of change of log-odds of the predicted risk.

2.6 Logistic Regression

Logistic regression (LR) is often used to investigate the relationship between discrete responses and explanatory variables and is widely accepted in the medical domain because it provides easy explanation/interpretation for predictions. Specifically, the log odds ratios describe the impact of the corresponding variables on the predicted results. Consider a dichotomous response variable such as CIH use, absent (0) or present (1), and a vector of variable(s) X that takes the value 1 if present and 0 if absent. The odds ratio is defined as the ratio of the odds for those with the variable=1 to the odds for those with variable=0. The formula and underlying assumptions for the odds ratio (e.g. sigmoid function) shares some aspects with the impact assessment above, and is given by:

$$ln\left(\frac{P}{1-P}\right) = a + bX$$

The probability *(P)* can be computed from the regression equation. So, if we know the regression equation, we can calculate the expected probability that $Y = 1$ for a given value of X.

$$P = \frac{e^{a+bx}}{1+e^{a+bx}}$$

We trained a logistic regression model as a comparison to our DNN/Impact Score analysis. The same source data was used as in DNN, but with an 80%-20% Train-Test split. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) solver was used.

2.7 Comparing DNN and Logistic Regression

To compare the DNN and Logistic Regression we use 3 measures between the log-odds ratio of the logistic regression and the impact score of the DNN. For an overall performance measure, we calculated the Area Under the Curve (AUC) of the Receiver Operating Characteristics. We also calculated the Pearson correlation to measure the amount of linear correlation, and Spearman correlation to compare the rank orders.

3. Results

Logistic regression and DNN with impact scores were performed on a dataset consisting of 19 variables from 9,761 patients for prediction of CIH use. The logistic regression required 83 iterations for convergence and finished with an AUC of 0.6805. The DNN impact scores are averages of 25 trainings of the DNN using different random seeds, with an average AUC of 0.7275. Details of the results are presented in Table 5, showing the log-odds ratio and 95% confidence interval for each variable of the logistic regression, and impact score with 95% confidence interval for each variable of the DNN. Correlation between the sets of results are shown in Table 6, with a Pearson correlation of 0.84 and Spearman correlation of 0.97.

To investigate the impact of variables that were not significant according to the LR model, we repeated the analysis with the non-significant variables excluded. Variables with non-significant p-values were Opioid Analgesics ($p>0.008$), Pain ($p>0.909$), Hypertension ($p>0.025$), BMI ($p>0.674$), Smoking History ($p>0.097$) and Hispanic ($p>0.088$). This resulted in increased

Table 5. Comparison of Logistic Regression to DNN Impact Score, ordered by log-odds ratio and impact score

Logistic Regression				DNN Impact Score			
	log-odds ratio	95% Conf Int			Impact score	95% Conf Int	
Drug Use Disorder	0.78	0.57	1.00	**Drug Use Disorder**	0.69	0.68	0.71
PTSD	0.73	0.59	0.86	**PTSD**	0.68	0.67	0.68
Alcohol Use Disorder	0.52	0.36	0.69	**Alcohol Use Disorder**	0.52	0.51	0.53
Gender (F)	0.50	0.34	0.67	**Bipolar Disorder**	0.49	0.48	0.50
MDD	0.47	0.30	0.65	**Gender (F)**	0.47	0.47	0.48
Bipolar Disorder	0.47	0.23	0.71	**Non-Opioid Analgesics**	0.47	0.46	0.47
Non-Opioid Analgesics	0.46	0.36	0.56	**MDD**	0.28	0.27	0.28
Black	0.41	0.22	0.60	**Anxiety Disorder**	0.27	0.26	0.28
Anxiety Disorder	0.28	0.12	0.44	**Black**	0.23	0.21	0.25
White	0.27	0.10	0.45	**BMI**	0.22	0.15	0.29
Hispanic	0.20	-0.03	0.44	**Opioid Analgesics**	0.13	0.12	0.13
Opioid Analgesics	0.14	0.04	0.24	**Age**	0.09	0.08	0.09
Smoking History	0.09	-0.02	0.19	**White**	0.08	0.07	0.10
Pain	0.01	-0.14	0.15	**Smoking History**	0.06	0.05	0.07
BMI	0.002	-0.01	0.01	**Hispanic**	0.03	0.00	0.05
Hypertension	-0.12	-0.23	-0.02	**Pain**	-0.07	-0.08	-0.06
CAD	-0.34	-0.50	-0.18	**Hypertension**	-0.17	-0.18	-0.17
Age	-1.35	-1.63	-1.06	**CAD**	-0.22	-0.22	-0.21
CCI	-1.75	-2.35	-1.15	**CCI**	-1.40	-1.46	-1.34
AUC = 0.6805				**Ave. AUC = 0.7275**			

Table 6. Correlations between the logistic regression log-odds ratios and the DNN impact scores.

Pearson Correlation (linear correlation)	**0.84**
Spearman Correlation (rank correlation)	**0.97**

Pearson correlation, but the other measures were largely unchanged (LR AUC = 0.6821, DNN AUC = 0.7278; Pearson correlation = 0.98; Spearman correlation = 0.96).

4. Discussion

We analyzed patient characteristics associated with the utilization of complementary and integrative health in patients with musculoskeletal disorders using a novel explainable deep learning approach and a traditional logistic regression approach. Both approaches identified Drug Use Disorder, PTSD, Alcohol Use Disorder, Gender (Female), MDD, Bipolar Disorder, Opioid and Non-Opioid Analgesic use, Race, and Anxiety Disorder as the characteristics associated with increased CIH use. CCI, CAD, and Hypertension were associated with decreased use of CIH.

Importantly, the results from the two approaches have some differences as well. Greater than average BMI, Smoking History, and Hispanic race showed a small association with increased CIH with the DNN model but were indeterminate with the logistic regression model due to its 95% confidence interval crossing 0. Conversely, Pain had a small association with decreased CIH in DNN but was not significant in logistic regression. Interestingly, these small

differences occurred in variables that had insignificant p-values in the LR model.

The general agreement between the DNN impact scores and the log odds ratios from LR reflect that both models were trained on the same data. Because DNN can have a variety of architectures and use different random seeds, multiple DNN models may be trained on a dataset. Combining the impact scores from multiple DNN models allows us to calculate a confidence interval for an impact score. Even though we cannot directly compare an impact score's confidence interval with the confidence interval of log odds ratio from the regression analysis, it does provide a range of potential values for the impact score in addition to the mean. The Pearson correlation and Spearman correlation show that the results are highly correlated both linearly and by rank order.

The goal of the analysis is NOT to predict CIH, but to understand the relationship between patient characteristics and CIH use. Since the decision to use CIH is often influenced by non-clinical characteristics such as provider or patient preference and availability of CIH services, we did not expect the logistic regression and DNN models to be able to predict CIH use based on only patient characteristics. Consistent with our expectation, the AUC for both logistic regression and the DNN impact scores were too low to be reliable.

The fact that the two different approaches arrive at similar but modestly different conclusions is intriguing. The novel DNN explanation method provides an alternative means to determine the effect of patient characteristics on CIH that does not depend on a priori assumptions and limitations of traditional statistical models. It can potentially detect novel and unexpected types of associations that would have to be decided upon in advance for traditional statistical models, partly due to the ability of DNNs to model non-linear relationships.

The points on which the two models agree imply that CIH is more likely to be used by those with a history of substance abuse and mental illness. This may be the result of physicians being more reluctant in prescribing opioids to those patients, resulting in more use of CIH. Also, worth noting is the stronger association of the use of non-opioid analgesics with CIH use than the use of opioid analgesics with CIH use.

An important limitation of this study is that CIH use is not completely captured by EHRs. On the other hand, based on our experience in CIH research, missing data of CIH use is more random than systematic. This limits the impact of the missing data on our analysis. Socio-economic status may be an important confounder and needs to be captured in follow up studies. In addition, temporal characteristics of the variables were not included in this analysis, such as length of chronic pain.

Future work will incorporate a more nuanced study of the relationships between opioid and other treatment options. In this study all CIH modalities were treated as one, however more knowledge is obtainable by differentiating the CIH modalities. In addition, patient and provider geographical location can be incorporated in order to discover variances by location, and also to incorporate socio-economic data stratified by location.

5. Conclusion

In this study we have demonstrated the practicality of a novel Impact Assessment method to interpret DNN models, for the purpose of exploring factors associated with utilization of CIH treatments among patients with musculoskeletal disorders. The DNN and logistic regression-based approaches arrived at similar but modestly different conclusions, while DNN does not depend on a priori assumptions and limitations of traditional statistical models, and is able to represent more complex, non-linear relationships.

6. References

[1] Martins SS, Sampson L, Cerda M, Galea S. Worldwide Prevalence and Trends in Unintentional Drug Overdose: A Systematic Review of the Literature. Am J Public Health. 2015;105(11):E29-E49.

[2] Centers for Disease Control and Prevention aNCfHS. Multiple cause of death 1999–2014 on CDC WONDER online database, released 2015. Data are from the Multiple Cause of Death Files, 1999–2014, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. 2016.

[3] Bohnert ASB, Ilgen MA, Galea S, McCarthy JF, Blow FC. Accidental Poisoning Mortality Among Patients in the Department of Veterans Affairs Health System. Med Care. 2011;49(4):393-6.

[4] Seal KH, Shi Y, Cohen G, Cohen BE, Maguen S, Krebs EE, et al. Association of mental health disorders with prescription opioids and high-risk opioid use in US veterans of Iraq and Afghanistan. JAMA. 2012;307(9):940-7.

[5] Strayer RJ, Motov SM, Nelson LS. Something for pain: Responsible opioid use in emergency medicine. Am J Emerg Med. 2017;35(2):337-41.

[6] Makary MA, Overton HN, Wang PQ. Overprescribing is major contributor to opioid crisis. Bmj-Brit Med J. 2017;359.

[7] Nalamachu SR, Shah B. Abuse of immediate-release opioids and current approaches to reduce misuse, abuse, and diversion. Postgrad Med. 2018:1-7.

[8] Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research. The National Academies Collection: Reports funded by National Institutes of Health. Washington (DC)2011.

[9] Taylor SL, Elwy AR. Complementary and alternative medicine for US veterans and active duty military personnel: promising steps to improve their health. Med Care. 2014;52(12 Suppl 5):S1-4.

[10] Denneson LM, Corson K, Dobscha SK. Complementary and alternative medicine use among veterans with chronic noncancer pain. J Rehabil Res Dev. 2011;48(9):1119-27.

[11] Baldwin CM, Long K, Kroesen K, Brooks AJ, Bell IR. A profile of military veterans in the southwestern United States who use complementary and alternative medicine: implications for integrated care. Arch Intern Med. 2002;162(15):1697-704.

[12] Chou R, Deyo R, Devine B, Hansen R, Sullivan S, Jarvik JG, et al. The Effectiveness and Risks of Long-Term Opioid Treatment of Chronic Pain. Evid Rep Technol Assess (Full Rep). 2014(218):1-219.

[13] Furlan AD, Giraldo M, Baskwill A, Irvin E, Imamura M. Massage for low-back pain. Cochrane Database Syst Rev. 2015(9):CD001929.

[14] Coeytaux RR, McDuffie J, Goode A, Cassel S, Porter WD, Sharma P, et al. Evidence Map of Yoga for High-Impact Conditions Affecting Veterans. VA Evidence-based Synthesis Program Reports. Washington (DC)2014.

[15] Cramer H, Lauche R, Haller H, Dobos G. A systematic review and meta-analysis of yoga for low back pain. Clin J Pain. 2013;29(5):450-60.

[16] Crawford C, Lee C, Buckenmaier C, 3rd, Schoomaker E, Petri R, Jonas W, et al. The current state of the science for active self-care complementary and integrative medicine therapies in the management of chronic pain symptoms: lessons learned, directions for the future. Pain Med. 2014;15 Suppl 1:S104-13.

[17] Furlan AD, Yazdi F, Tsertsvadze A, Gross A, Van Tulder M, Santaguida L, et al. A systematic review and meta-analysis of efficacy, cost-effectiveness, and safety of selected complementary and alternative medicine for neck and low-back pain. Evid Based Complement Alternat Med. 2012;2012:953139.

[18] Hempel S, Taylor SL, Solloway MR, Miake-Lye IM, Beroes JM, Shanman R, et al. Evidence Map of Acupuncture. VA Evidence-based Synthesis Program Reports. Washington (DC)2014.

[19] Peterson K, Anderson J, Ferguson L, Mackey K. Evidence Brief: The Comparative Effectiveness of Selected Complementary and Integrative Health (CIH) Interventions for Preventing or Reducing Opioid Use in Adults with Chronic Neck, Low Back, and Large Joint Pain. VA Evidence Synthesis Program Evidence Briefs. VA Evidence Synthesis Program Reports. Washington (DC)2011.

[20] Kizer KW, Dudley RA. Extreme makeover: Transformation of the veterans health care system. Annu Rev Public Health. 2009;30:313-39.

[21] Goulet JL, Erdos J, Kancir S, Levin FL, Wright SM, Daniels SM, et al. Measuring performance directly using the Veterans Health Administration electronic medical record - A comparison with External Peer Review. Med Care. 2007;45(1):73-9.

[22] Kizer KW, Demakis JG, Feussner JR. Reinventing VA health care: systematizing quality improvement and quality innovation. Med Care. 2000;38(6 Suppl 1):I7-16.

[23] (HAIG) VHAIG. Complementary and Alternative Medicine. 2011.

[24] Redd D, Kuang J, Zeng-Treitler Q. Differences in nationwide cohorts of acupuncture users identified using structured and free text medical records. AMIA Annu Symp Proc. 2014;2014:1002-9.

[25] Goulet JL, Kerns RD, Bair M, Becker WC, Brennan P, Burgess DJ, et al. The musculoskeletal diagnosis cohort: examining pain and pain care among veterans. Pain. 2016;157(8):1696-703.

[26] Redd D, Kuang J, Mohanty A, Bray BE, Zeng-Treitler Q. Regular expression-based learning for METS value extraction. AMIA Summits on Translational Science Proceedings. 2016;2016:213.

[27] Redd D, Frech TM, Murtaugh MA, Rhiannon J, Zeng QT. Informatics can identify systemic sclerosis (SSc) patients at risk for scleroderma renal crisis. Computers in biology and medicine. 2014 Oct 1;53:203-5.

[28] Redd D, Shaoa Y, Yang J, Divita G, Zeng-Treitler Q. Automated Learning of Temporal Expressions. Studies in health technology and informatics. 2015;216:639-42.

[29] Murtaugh MA, Gibson BS, Redd D, Zeng-Treitler Q. Regular expression-based learning to extract bodyweight values from clinical notes. Journal of biomedical informatics. 2015 Apr 1;54:186-90.

[30] Ackerman CJ, Turkoski B. Using guided imagery to reduce pain and anxiety. Home Healthcare Now. 2000 Sep 1;18(8):524-30.

[31] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining San Francisco, California, USA 2016. p. 1135-44.

[32] Shao Y, Cheng Y, Shah R, Weir C, Bray B, Zeng-Treitler Q. Shedding Light on the Black Box: Explaining Deep Neural Network Prediction of Clinical Outcome. ICHI 2019 : International Conference on Health Informatics; Rome, Italy2019.