How is the Doctor Feeling? ICU Provider Sentiment is Associated with Diagnostic Imaging Utilization

Mohammad M. Ghassemi†, Tuka Alhanai†, Jesse D. Raffa,
Roger G. Mark, Shamim Nemati and Falgun H. Chokshi

***Abstract*— The judgment of intensive care unit (ICU) providers is difficult to measure using conventional structured electronic medical record (EMR) data. However, provider sentiment may be a proxy for such judgment. Utilizing 10 years of EMR data, this study evaluates the association between provider sentiment and diagnostic imaging utilization. We extracted daily positive / negative sentiment scores of written provider notes, and used a Poisson regression to estimate sentiment association with the total number of daily imaging reports. After adjusting for confounding factors, we found that (1) negative sentiment was associated with increased imaging utilization ($p < 0.01$), (2) sentiment's association was most pronounced at the beginning of the ICU stay ($p < 0.01$), and (3) the presence of any form of sentiment increased diagnostic imaging utilization up to a critical threshold ($p < 0.01$). Our results indicate that provider sentiment may clarify currently unexplained variance in resource utilization and clinical practice.**

I. Introduction

As the United States (US) national healthcare expenditure continues to rise [1], the use of high-cost medical resources has come under increased scrutiny [2]. The US Health and Human Services is already mandating more judicious use of high cost medical resources, motivating investigation into the historical drivers of resource utilization, and the development of more specific criteria for justifying the utilization of high-cost resources [3]. One prominent example of a high cost medical resource is radiological diagnostic imaging, which accounts for nearly 10% of US health-care expenditures [4]. Diagnostic imaging (hereafter, imaging) is frequently cited as a high-cost medical resource in need of better defined appropriateness criteria and this year, in an effort to curtail imaging costs, the US Protecting Access to Medicare Act will mandate the use of clinical decision support tools [3].

†Both authors contributed equally to this work.

M.M. Ghassemi is with Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology,Boston, MA 02139 USA. Phone: 617-599-6010. Email: ghassemi@mit.edu

T. Alhanai is with Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology,Boston, MA 02139 USA. Email: tuka@mit.edu

J.D. Raffa is with the Institute of Medical Engineering and Science, Massachusetts Institute of Technology,Boston, MA 02139 USA. Email: jraffa@mit.edu

Roger G. Mark is with the Institute of Medical Engineering and Science, Massachusetts Institute of Technology,Boston, MA 02139 USA. Email: rgmark@mit.edu

S. Nemati is with the Department of Biomedical Informatics, Emory University, Atlanta, GA 30322 USA. Email: shamim.nemati@emory.edu

F.H. Chokshi is with the Department of Radiology and Imaging Services, Emory University School of Medicine, Atlanta, GA 30322 USA. Email: falgun.chokshi@emory.edu

Several studies over the last decade have compared the diagnostic yield and cost-effectiveness of imaging utilization across a variety of cohort sizes (from small single center studies [5], [6], [7], [8], [9], to large Medicare claims-mining [10], [11]) and patient conditions (including ankle fractures [12], incidental lung nodules [13], altered mental status [5], and head injury [14]). Most of these utilization studies are performed retrospectively, using structured data from the electronic medical record (EMR) or administrative claims databases (e.g. Medicare) to identify ordered examinations, diagnoses, and sociodemographic information that are associated with utilization rates. By relying exclusively on structured data, many retrospective analyses fail to capture the complete clinical context considered by health-care providers when ordering imaging exams [15], [16].

At the bedside, provider judgment reflects observations that may or may not be entirely reflected in structured medical data. It follows that an estimate of this judgment may help explain a previously unknown component of the variance in utilization patterns during treatment, and consequently, healthcare costs.

In this paper, we investigate the utility of the sentiment in electronic provider notes as a proxy of this provider judgment [17]. Specifically, we investigate how provider sentiment is

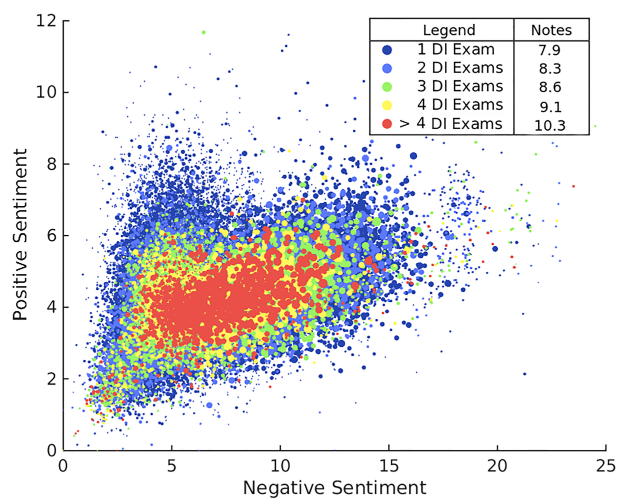

Fig. 1. **Sentiment in Medical Notes.** Each point represents a patient's day in the ICU. Colors represent the number of radiological exams received (see legend), while the size of each point indicates the number of provider notes used to compute the sentiment. The average number of notes for each radiological exam level is shown in the figure legend.

associated with diagnostic imaging utilization, after adjusting for the effects of severity of illness, comorbidities, and other factors.

Sentiment analysis is a branch of natural language processing (NLP) that combines text analysis and computational linguistics to assess the emotion or polarity of a piece of text (positive, negative, or neutral) [18]. Sentiment analysis has been used widely in non-healthcare settings, such as social media [19] and newspaper publications [20] to identify and extract text-based sentiment. In the last few years this approach has also found application in the evaluation of health-related topics including health and happiness [21], health care satisfaction [22], and health care reform [23]. Even more recently, direct analysis of sentiment within medical records has emerged as a topic of research. Applying word embedding, Ghassemi *et al.* [17] explored the relationship between provider sentiment, patient demographics, and mortality using sentiment analysis of structured and unstructured EMR data of intensive care unit (ICU) patients. McCoy *et al*. employed a sentiment estimate to examine associations between sentiment, readmission, and mortality risk using hospital discharge notes alone [24]. However, no publications we are aware of have investigated the association between provider sentiment and resource utilization patterns.

II. Objective

This investigation aimed to answer the following two questions: (1) what was the relationship between ICU provider sentiment and diagnostic imaging utilization? (2) Was this relationship consistent over the course of ICU length of stay, or did it change over time?

III. Methods

A. Data Sources and Settings

All data for this study were extracted from the publicly available Medical Information Mart for Intensive Care (MIMIC-III) database, which contains the structured and unstructured EMR data of over fifty thousand patients from the Beth Israel Deaconess Medical Center in Boston MA, from 2001 - 2012 [25]. We were interested in understanding the association between the sentiment of provider free-text notes in MIMIC-III, and the total number of daily diagnostic imaging examinations per patient.

B. Study Variables

The outcome of interest was the total number of daily radiology reports for each ICU patient as a surrogate for the total number of imaging exams. We extracted a set of continuous and categorical features that we suspected, based on clinical experience, might confound the relationship between provider sentiment and the number of daily imaging exams. The features included: patient age, the Sequential Organ Failure Assessment Score (SOFA), the Elixhauser comorbidity index [26] and the Oxford Acute Severity of Illness Score (OASIS) [25], gender (with female being the reference group) and ethnicity (white, black, Hispanic and other, with white as the reference group). We also included dichotomous indicators for the following conditions: obesity, human immunodeficiency virus infection (HIV), metastatic cancer diagnosis, diabetes and ICU type (with surgical coded as one).

C. Eligibility Criteria and Study Size

The MIMIC-III database contains notes of several distinct categories. For this analysis we only considered provider notes from the first five days of patient ICU stay which were of the following types: Consult, General, Nursing, Nutrition, Pharmacy, Physician, Rehabilitation Services and Respiratory. We excluded the notes of all neonatal patients and those missing any of the covariates described above. The exclusion criteria reduced the number of distinct notes from 697,718 to 283,950, the number of distinct ICU stays from 52,420 to 18,607 and the number of distinct days of data from 129,624 to 45,728. 76% of the excluded patients were neonates, while 23% were excluded due to missing covariates. Figure 2 illustrates the number of ICU stays and notes removed from our dataset for each step of our exclusion criteria.

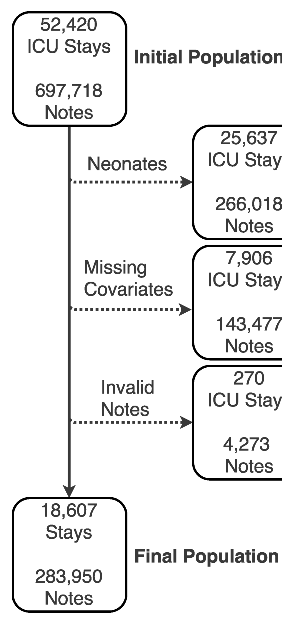

Fig. 2. **Study Population.** The effects of our exclusion criteria on the initial sample.

D. Sentiment Features

For every word w in every ICU provider document d, we extracted a corresponding measure of positive sentiment $p_{d,w}$, and negative sentiment $n_{d,w}$, according to the first sense of the word as defined by SentiWordNet [27]. SentiWordNet is an open-source tool that provides both the sentimental polarity (positive or negative) and sentimental magnitude (ranging from 0 to 1) of commonly used words in the English language. From these scores we estimated document level positive and negative sentiment, p_d and n_d as the average sentiment scores of all words in a given document:

$$p_d = \frac{1}{M_d}\sum_{w=1}^{M_d} p_{d,w} \quad ; \quad n_d = \frac{1}{M_d}\sum_{w=1}^{M_d} n_{d,w}$$

where M_d represents the total number of SentiWordNet words in a document. Finally, a daily patient sentiment score was computed as the average of all document-level scores on a given day. This score was multiplied by 100 to aid in its interpretation during later modeling. Hence, the word, and document sentiment scores are continuous and in the range of 0 to 1, while the daily sentiment scores used in all the analysis of this work are continuous and in the range of 0 to 100.

E. Statistical Methods

A multivariable Poisson regression was used to model the relationship between our features and the number of daily imaging exams while adjusting for potential confounders. Generalized estimating equations (GEE) with a Markov correlation structure [28] were used to fit the models. GEE is useful for our purposes as it accounts for the correlated nature of the observations within each patient, across multiple ICU days and allows for patients to serve as their own controls. For patient j, on day t, we model our outcome as $Y_{jt} \sim Possion(\mu_{jt})$, with a log-linked mean μ_{jt}, which can be computed as a function of our time-varying features $\mathbf{x}_{jt}$:

$$log(E[y_{jt}|x_{jt}]) = log(\mu_{jt}) = \mathbf{x}_{jt}\hat{\beta}$$

Where we estimate the coefficients, $\hat{\beta}$, using GEE. It can be somewhat difficult to interpret a given coefficient (β_k) of a Poisson regression in its raw form:

$$\beta_k = log(E[y|x_k + 1, \mathbf{x}_{\hat{k}}]) - log(E[y|x_k, \mathbf{x}_{\hat{k}}])$$

Where x_k represents the feature of interest, and $\mathbf{x}_{\hat{k}}$ represents all other features in $\mathbf{X}$. To aid in the interpretation of our model, we report a commonly used transformation of these coefficients which reflect the multiplicative effects of a one unit change in the corresponding feature x_k on the outcome y:

$$\frac{E[y|x_k + 1, \mathbf{x}_{\hat{k}}]}{E[y|x_k, \mathbf{x}_{\hat{k}}]} = e^{\beta_k}$$

This can be interpreted as the ratio of the rates of the outcome, y, estimated when changing one unit of x_k, while keeping all other variables the same.

F. Study Design

The investigation began by creating descriptive summaries and visualizations of the population data in both tabular and graphical formats (Presented in Table I and Figure 1). Visual inspection motivated three targeted analyses.

The first analysis treated sentiment as a continuous quantity and assumed a linear relationship between sentiment and the log rate of diagnostic imaging utilization (Presented in Table II). The second analysis was nearly identical to the first, but included quadratic terms for the sentiment features in the model (Presented in Appendix Table III).In the last analysis, we evaluated the modifying effect of the day of ICU stay on the sentiment features associations found in the first two parts of the analysis by including day **x** sentiment interaction terms for each of the first five days (Presented in Figures 3, 4 and Appendix 5). This allowed us to consider cases where the strength of the association between the sentiment variables and the rate of image utilization varies, as the patient stays longer in the ICU.

Statistical significance of the individual coefficients or groups of coefficients was assessed using Wald tests. In addition to assessing the statistical significance, the QICu [29] of the models were computed with and without the sentiment features to further evaluate any evidence of improved model fit facilitated by the sentiment variables in predicting the rate of imaging utilization. QICu is similar to the Akaike Information Criteria, but modified for models fit with GEE.

IV. RESULTS

A. Descriptive Data of Participants and Outcomes

Table I shows summary statistics for the sample of ICU days partitioned by the number of imaging exams. Figure 1 (first page) presents a bubble plot that simultaneously represents the sample's daily sentiment, number of imaging exam reports, and total number of written ICU provider notes. The figure reveals visually that the number of imaging exams (represented by color) decreases as positive sentiment (y-axis) increases. The plot also reveals a potential curvilinear relationship between sentiment and imaging utilization. That is, increased negative sentiment (x-axis) is associated with an increased number of imaging exams (color), up until a negative sentiment value of approximately ten, after which

TABLE I

SUMMARY STATISTICS. EXTRACTED FEATURES FOR THE PATIENT POPULATION, PARTITIONED BY THE NUMBER OF DAILY IMAGING EXAMS. OASIS: OXFORD ACUTE SEVERITY OF ILLNESS SCORE. SOFA: SEQUENTIAL ORGAN FAILURE ASSESSMENT.

EXAMS PER DAY	**1**	**2**	**3**	**4**
Sample Size (Days)	25,455	11,867	4,829	3,548
CONTINOUS FEATURES (Mean [Standard Deviation])				
Age	65.1 (16)	64.1 (17)	63.2 (17)	61.6 (17.5)
OASIS	30.6 (9)	30.9 (9)	31.5 (9)	32.2 (9)
Elixhauser	4.0 (2)	3.9 (2)	3.9 (2.1)	3.9 (2.1)
SOFA	3.7(3)	4.0 (3)	4.4 (3.2)	4.8 (3.2)
CATEGORICAL FEATURES (%)				
Diabetes	29.3	27.4	25.5	25.2
HIV infection	1.2	1.3	1.1	1.1
Ethnicity - Hispanic	3.6	3.7	3.15	4.2
Ethnicity - Black	7.3	7.1	6.8	7.9
Ethnicity - Other	15.6	15.9	16.2	14.2
Gender (Male)	57.0	58.3	59.0	60.1
Cancer	5.6	6.1	6.2	5.8
Obesity	6.7	6.6	6.2	6.8
ICU Type (Surgical)	17.2	19.8	22.5	22.1
SENTIMENT FEATURES (Mean [Standard Deviation])				
Negative	6.3 (2.6)	6.4 (2.4)	6.4 (2.3)	6.5 (2.2)
Positive	4.4 (1.2)	4.3 (0.9)	4.3 (0.8)	4.2 (0.7)

additional negative sentiment is associated with fewer exams. This would suggest that the maximum rate of image utilization might occur at some intermediate levels of sentiment.

B. Main Results

1) Effects of Linear Sentiment: The results of the Poisson regression using linear sentiment features are shown in Table II. The following confounding features exhibited statistically significant association with higher imaging utilization: OASIS, SOFA, gender (male), and ICU type (surgical) or lower utilization: older age, HIV status, diabetes and ethnicity (other). Of particular interest is the statistical significance of the positive and negative sentiment features (last two rows of Table II), which provides evidence that provider sentiment is associated with the number of imaging exams even after adjusting for the effects of the selected confounders. The relative rate ratio (ϵ^{β}) shown in Table II explains the multiplicative effects of a one-unit increase in sentiment, on the total number of daily imaging exams. Under this model we found that a 1% increase in positive sentiment was associated with a 4% decrease in the rate of daily imaging exams while a 1% increase in negative sentiment is associated with a 1% increase in the rate of daily imaging exams.

2) Imaging Utilization over Time: Figure 3 illustrates the temporal evolution of sentiment's linear association with imaging utilization. Specifically, the figure illustrates an estimate of the relative rate ratios and 95% confidence intervals of positive and negative sentiment (allowing one term for each day of treatment). The trend in Figure 1 indicates that provider sentiment is more strongly associated with imaging utilization on the first day of ICU stay, and grows weaker over subsequent days of care. Indeed, by the fourth and fifth days of care, the association between negative

TABLE II

RESULTS (LINEAR). MULTIVARIABLE POISSON REGRESSION MODEL WITH FEATURES SORTED BY THE STRENGTH OF THE COEFFICIENT, β . SE: STANDARD ERROR. CI: CONFIDENCE INTERVAL. OASIS: OXFORD ACUTE SEVERITY OF ILLNESS SCORE. SOFA: SEQUENTIAL ORGAN FAILURE ASSESSMENT. ETHN.: ETHNICITY. *RELATIVE TO WHITE

	β	**Odds Ratio (ϵ^{β})**	**CI (95%)**	**p-value**
CONTINUOUS FEATURES				
Age (per year)	-2.2E-3	0.99	0.9896 - 0.9904	<0.001
OASIS	2.2E-3	1.002	1.0012 - 1.0028	<0.001
Elixhauser	5.5E-4	1.00	0.0063 - 1.0037	0.74
SOFA	1.8E-2	1.02	1.0178 - 1.0222	<0.001
CATEGORICAL FEATURES				
HIV infection	-6.4E-2	0.94	0.8890 - 0.9910	0.01
Diabetes	-3.4E-2	0.97	0.9547 - 0.9853	<0.001
Ethn. - Other*	-2.1E-2	0.98	0.9626 - 0.9974	0.02
Ethn. - Hispanic*	-7.5E-3	0.99	0.9547 - 1.0253	0.67
Ethn. - Black*	-4.7E-3	0.99	0.9651 - 1.0174	0.71
Obesity	-8.1E-3	0.99	0.09626 - 1.0174	0.56
Gender (Male)	1.7E-2	1.02	1.0069 - 1.0331	0.01
Cancer	2.2E-2	1.02	0.9947 - 1.0453	0.09
ICU Type (Surgical)	7.4E-2	1.08	1.0637 - 1.0963	<0.001
SENTIMENT FEATURES				
Positive	-4.4E-2	0.96	0.9580 - 0.9620	<0.001
Negative	9.4E-3	1.01	1.0061 - 1.0139	<0.001

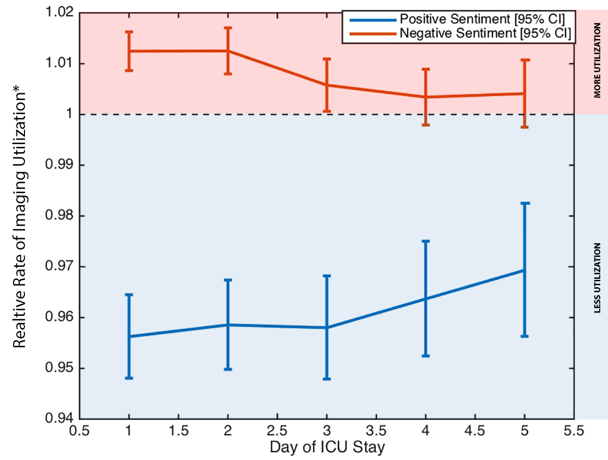

Fig. 3. **Imaging Utilization over Time.** Changes in image utilization as a function of daily sentiment over the first 5 days of ICU stay. Curves represent the value of the relative rate of image utilization per unit increase in daily sentiment (y-axis) by ICU day (x-axis), while adjusting for the confounders listed in II. The model was trained using all features shown in Table II, in addition to separate day effects and day **x** sentiment interactions.

sentiment (red line) and imaging utilization is no longer statistically significant. Importantly, the temporal trends in both the positive and negative sentiment were found to be statistically significant (p = 0.04) when assessing the statistical significance of the day **x** sentiment interactions. These findings suggest that providers may rely more heavily on personal observation and judgment (which is reflected in their sentiment) in the early days of treatment, when less objective clinical information is known about the patient.

3) Effects of Quadratic Sentiment: In Figure 1, we observed visual evidence of a potential quadratic relationship between sentiment and imaging utilization. To study this relationship more formally, an extension of the model described in Table II was generated, including quadratic terms for both sentiment scores. Model coefficients are presented in Table III in the Appendix. The quadratic terms for both sentiment features were statistically significant ($p < 0.001$) and their inclusion in the model was merited according to the QICu metric. This suggests that the presence of any form of sentiment may increase imaging utilization up until a critical threshold, beyond which utilization is reduced. This phenomenon may reflect the reluctance of ICU providers to order additional imaging exams on patients that may be either doing very well clinically, or so poorly that death is imminent.

4) Imaging Utilization as a Function of Sentiment: Figure 4 illustrates the relationship between the relative rate of imaging exams and an increase in negative sentiment, where this relationship is permitted to vary over the ICU stay day (holding all other variables fixed). The strength of the effect of changes in negative sentiment (x-axis) on relative imaging utilization rate (y-axis) seems to be particularly strong in day one (blue line), after which the effect is attenuated in days two through five (See Figure 5 in the appendix for a similar illustration using positive sentiment). Importantly, in Figure 4, the direction of association between negative sentiment and imaging utilization reverses at different sentiment thresh-

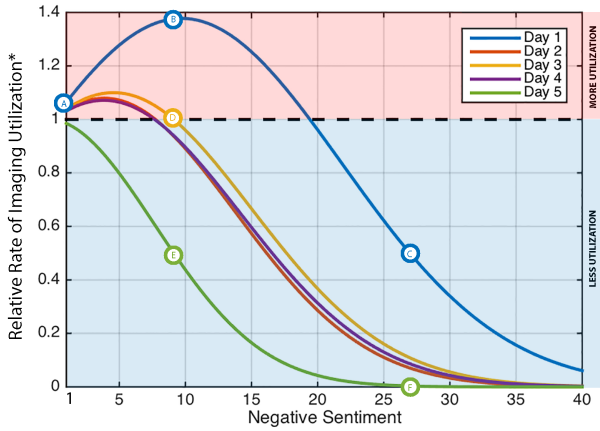

Fig. 4. **Imaging Utilization as a Function of Sentiment.** The association between negative sentiment (y-axis) and imaging utilization over time for the model with linear and quadratic sentiment terms. Curves represent the estimated relative rate of image utilization per unit increase in negative sentiment, estimated separately for each ICU day. A value of one is interpreted as no effect. The model was trained using all features shown in Table III (Appendix) in addition to separate day effects, as well as both day **x** sentiment and day **x** sentiment^2 interactions. See the last paragraph of the results section for an explanation of points A-F.

olds, on different days. On day one (blue line), the initial effects of negative sentiment are associated with increased imaging utilization (see point A, on blue line). The strength of the association between negative sentiment and increased imaging utilization continues to grow, up until a negative sentiment value of approximately ten (see point B, on blue line). While values of negative sentiment between ten and twenty continue to be associated with increased imaging utilization, the strength of this association weakens, eventually causing negative sentiment to be associated with significantly less imaging utilization (see point C, on blue line). There is a similar general relationship between negative sentiment and utilization on day three (orange line). However, the strength of the association is significantly less pronounced, with a maximum imaging utilization effect of 1.1 on day three, compared to 1.35 on day one. Furthermore, while a negative sentiment value of ten had a strong association with utilization on day one (see point B, on blue line), the same level of negative sentiment has no effect on day 3 (see point D, on orange line), and had a negative association on day five (see point E, on green line). Indeed, by the fifth day of treatment (green line), negative sentiment only serves to decrease utilization, and more rapidly reaches the point of zero utilization (see point F, on green line) than any other day. A Wald test for the statistical significance of the day **x** sentiment interactions was statistically significant ($p = 0.04$), and the QICu was the lowest of any model we fit.

V. Discussion

A. Key Results

This study presented a novel investigation of the association between ICU provider sentiment and imaging utilization. While many studies have investigated imaging utilization using structured data, none to date have managed to isolate the specific effects of provider judgment or observation on imaging utilization as we have shown in this study. Furthermore, prospective study of clinical intent is, as yet, impractical because requiring clinicians to report their intent may artificially alter their utilization behavior. The results of this paper may be summarized in the following four points: (1) there was a statistically significant association between ICU provider sentiment and imaging utilization; (2) in general, positive sentiment was associated with less imaging exams performed, while negative sentiment was associated with more exams; (3) provider sentiment that was strongly positive, or strongly negative, was associated with decreased imaging utilization; and (4) the association between sentiment and imaging utilization was strongest on the first day of ICU stay and grew gradually weaker over time.

B. Significance

This is the first study to investigate how provider sentiment is associated with diagnostic imaging utilization, after adjusting for the effects of severity of illness, age, comorbidities, and length of ICU stay. Although we focused on imaging utilization, the implications of this work extend well beyond imaging. As a proxy for decision-making judgment, sentiment analysis has significant potential to assess utilization of other resources, such as laboratory tests, use of ancillary services, and discharge/transfer decisions. It may also help explain variance in other areas of clinical practice, including medication dosing [30], and decisions to withdraw care [31].

C. Limitations

One potential limitation of this study is that the word-level sentiment scores provided by SentiWordNet were not designed to account for the complexities of provider notes or specialized medical language. SentiWordNet does not incorporate established medical lexicon and ontology sources, such as the Unified Medical Language System (UMLS [32]), Systematized Nomenclature of Medicine Clinical Terms (SNOWMED CT [33]), and Disease Ontology [34]. While this may be interpreted as a weakness of the presented approach, it may also be interpreted as a strength: we were not interested in the sentiment of the medical terms, but rather, the aggregate sentiment of the individual care providers, a dimension that is orthogonal to the clinical and physiological factors reflected in specialized clinical language of the notes.

D. Interpretation

Importantly, this work demonstrates that even the simplest estimates of ICU provider sentiment may have both statistically and clinically significant associations with imaging utilization. Provider sentiment showed independent effects on imaging utilization after controlling for patient sociodemographic factors (i.e., age, gender, race), comorbidities, severity of illness, and time in the ICU. We conclude that our measure of provider sentiment reflects something about the observations and judgment of the care provider that is not captured by the structured medical data alone. This additional clinical dimension has been missing in studies that have evaluated imaging utilization in high-cost health care settings [10], [5], [6], [7], [8], [11], [9]. In the current healthcare environment of increased scrutiny of high cost resources, the novel measure of provider judgment proposed in this study could better help calibrate clinical decision support systems to include the effects of clinical sentiment as a proxy of judgment.

Our results provide strong motivation for further investigation of the association between provider sentiment and patterns of resource use, namely imaging utilization. Indeed, our results may serve as a valuable baseline as more sophisticated techniques are developed or deployed (e.g. deep learning) [35]. With further development, investigators may begin to take steps towards providing actionable adjustments to ordering practices on the basis of sentiment analysis. It should be possible to design a decision support tool that provides additional input for high-cost decisions where there is a large difference between the expected and observed sentiment of the providers, or where multiple providers have conflicting sentiment about the same patient.

ACKNOWLEDGEMENTS

M.M. Ghassemi would like to acknowledge the Salerno foundation, The NIH Neuroimaging Training Grant (NTP -T32 EB 001680), the Advanced Neuroimaging Training Grant (AMNTP T90 DA 22759). T. Alhanai acknowledges the Abu Dhabi Education Council for their support. F.H. Chokshi would like to acknowledge the Association of University Radiologists (AUR) General Electric Radiology Research Academic Fellowship (GERRAF). S. Nemati would like to acknowledge the support of the NIH early career development award in biomedical big data science (K01ES025445). R.G. Mark, and J.D. Raffa would like to acknowledge the support of the NIH Critical Care Informatics Grant (R01-EB017205) and the NIH Research Resource for Complex Physiologic Signals Grant (R01-GM104987).

REFERENCES

[1] A. B. Martin, M. Hartman, J. Benson, A. Catlin, N. H. E. A. Team *et al.*, "National health spending in 2014: Faster growth driven by coverage expansion and prescription drug spending," *Health Affairs*, vol. 35, no. 1, pp. 150–160, 2016.

[2] B. Chang, J. Lorenzo, and A. Macario, "Examining health care costs: Opportunities to provide value in the intensive care unit," *Anesthesiology clinics*, vol. 33, no. 4, pp. 753–770, 2015.

[3] S. M. Burwell, "Setting value-based payment goalshhs efforts to improve us health care," *N Engl J Med*, vol. 372, no. 10, pp. 897–9, 2015.

[4] W. Howell. (2014) Imaging utilization trends and reimbursement. [Online]. Available: www.diagnosticimaging.com/reimbursement/imaging-utilization-trends-and-reimbursement

[5] F. H. Chokshi, G. Sadigh, W. Carpenter, J. Kang, R. Duszak Jr, and F. Khosa, "Altered mental status in icu patients: Diagnostic yield of noncontrast head ct for abnormal and communicable findings," *Critical Care Medicine*, vol. 44, no. 12, pp. e1180–e1185, 2016.

[6] N. Desai, J. Kang, and F. Chokshi, "Screening ct angiography for pediatric blunt cerebrovascular injury with emphasis on the cervical seatbelt sign," *American Journal of Neuroradiology*, vol. 35, no. 9, pp. 1836–1840, 2014.

[7] G. Moulin-Romsee, A. Maes, D. Silverman, L. Mortelmans, and K. Van Laere, "Cost-effectiveness of 18f-fluorodeoxyglucose positron emission tomography in the assessment of early dementia from a belgian and european perspective," *European journal of neurology*, vol. 12, no. 4, pp. 254–263, 2005.

[8] D. M. Panczykowski, N. D. Tomycz, and D. O. Okonkwo, "Comparative effectiveness of using computed tomography alone to exclude cervical spine injuries in obtunded or intubated patients: meta-analysis of 14,327 patients with blunt trauma: A review," *Journal of neurosurgery*, vol. 115, no. 3, pp. 541–549, 2011.

[9] P. Sanelli, A. Pandya, A. Segal, A. Gupta, S. Hurtado-Rua, J. Ivanidze, K. Kesavabhotla, D. Mir, A. Mushlin, and M. Hunink, "Cost-effectiveness of ct angiography and perfusion imaging for delayed cerebral ischemia and vasospasm in aneurysmal subarachnoid hemorrhage," *American Journal of Neuroradiology*, vol. 35, no. 9, pp. 1714–1720, 2014.

[10] F. H. Chokshi, D. R. Hughes, J. M. Wang, M. E. Mullins, C. M. Hawkins, and R. Duszak, "Diagnostic radiology resident and fellow workloads: a 12-year longitudinal trend analysis using national medicare aggregate claims data," *Journal of the American College of Radiology*, vol. 12, no. 7, pp. 664–669, 2015.

[11] A. M. Prabhakar, A. S. Misono, J. Hemingway, D. R. Hughes, and R. Duszak, "Medicare utilization of vascular ultrasound from 1998 to 2013: Continued growth in both radiologist and nonradiologist imaging," *Journal of the American College of Radiology*, 2015.

[12] P. C. Silveira, I. K. Ip, S. Sumption, A. S. Raja, S. Tajmir, and R. Khorasani, "Impact of a clinical decision support tool on adherence to the ottawa ankle rules," *The American journal of emergency medicine*, 2015.

[13] M. T. Lu, D. A. Rosman, C. C. Wu, M. D. Gilman, H. B. Harvey, D. A. Gervais, T. K. Alkasab, J.-A. O. Shepard, G. W. Boland, and P. V. Pandharipande, "Radiologist point-of-care clinical decision support and adherence to guidelines for incidental lung nodules," *Journal of the American College of Radiology*, vol. 13, no. 2, pp. 156–162, 2016.

[14] E. R. Melnick, J. Keegan, and R. A. Taylor, "Redefining overuse to include costs: A decision analysis for computed tomography in minor head injury," *The Joint Commission Journal on Quality and Patient Safety*, vol. 41, no. 7, pp. 313–313, 2015.

[15] J. M. Madden, M. D. Lakoma, D. Rusinak, C. Y. Lu, and S. B. Soumerai, "Missing clinical and behavioral health data in a large electronic health record (ehr) system," *Journal of the American Medical Informatics Association*, vol. 23, no. 6, pp. 1143–1149, 2016.

[16] M. S. Ridgely and M. D. Greenberg, "Too many alerts, too much liability: sorting through the malpractice implications of drug-drug interaction clinical decision support," . *Louis UJ Health L. & Pol'y*, vol. 5, p. 257, 2011.

[17] M. M. Ghassemi, R. G. Mark, and S. Nemati, "A visualization of evolving clinical sentiment using vector representations of clinical notes," in *Computing in Cardiology Conference (CinC), 2015*, Sept 2015, pp. 629–632.

[18] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*. Association for Computational Linguistics, 2002, pp. 79–86.

[19] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič, "The effects of twitter sentiment on stock price returns," *PloS one*, vol. 10, no. 9, p. e0138441, 2015.

[20] Y. Lei, J. A. Pereira, S. Quach, J. A. Bettinger, J. C. Kwong, K. Corace, G. Garber, Y. Feinberg, M. Guay, P. H. A. of Canada/Canadian Institutes of Health Research Influenza Research Network (PCIRN) Program Delivery, E. Group *et al.*, "Examining perceptions about mandatory influenza vaccination of healthcare workers through online comments on news stories," *PloS one*, vol. 10, no. 6, p. e0129993, 2015.

[21] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Donaldson, "Use of sentiment analysis for capturing patient experience from free-text comments posted online," *Journal of medical Internet research*, vol. 15, no. 11, p. e239, 2013.

[22] L. Mitchell, M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth, "The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place," *PloS one*, vol. 8, no. 5, p. e64417, 2013.

[23] D. King, D. Ramirez-Cano, F. Greaves, I. Vlaev, S. Beales, and A. Darzi, "Twitter and the health reforms in the english national health service," *Health policy*, vol. 110, no. 2, pp. 291–297, 2013.

[24] T. H. McCoy, V. M. Castro, A. Cagan, A. M. Roberson, I. S. Kohane, and R. H. Perlis, "Sentiment measured in hospital discharge notes is associated with readmission and mortality risk: An electronic health record study," *PloS one*, vol. 10, no. 8, p. e0136341, 2015.

[25] A. E. Johnson, A. A. Kramer, and G. D. Clifford, "A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy*," *Critical care medicine*, vol. 41, no. 7, pp. 1711–1718, 2013.

[26] A. Elixhauser, C. Steiner, D. R. Harris, and R. M. Coffey, "Comorbidity measures for use with administrative data," *Medical care*, vol. 36, no. 1, pp. 8–27, 1998.

[27] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining." in *LREC*, vol. 10, 2010, pp. 2200–2204.

[28] J. Shults and S. J. Ratcliffe, "Geeqbox: A matlab toolbox for generalized estimating equations and quasi-least squares," *Journal of Statistical Software*, vol. 25, no. 14, 2008.

[29] W. Pan, "Akaike's information criterion in generalized estimating equations," *Biometrics*, vol. 57, no. 1, pp. 120–125, 2001.

[30] M. M. Ghassemi, S. E. Richter, I. M. Eche, T. W. Chen, J. Danziger, and L. A. Celi, "A data-driven approach to optimized medication dosing: a focus on heparin," *Intensive care medicine*, vol. 40, no. 9, pp. 1332–1339, 2014.

[31] R. Pirracchio, M. L. Petersen, M. Carone, M. R. Rigon, S. Chevret, and M. J. van der Laan, "Mortality prediction in intensive care units with the super icu learner algorithm (sicula): a population-based study," *The Lancet Respiratory Medicine*, vol. 3, no. 1, pp. 42–52, 2015.

[32] C. Lindberg, "The unified medical language system (umls) of the national library of medicine." *Journal (American Medical Record Association)*, vol. 61, no. 5, p. 40, 1990.

[33] M. Q. Stearns, C. Price, K. A. Spackman, and A. Y. Wang, "Snomed clinical terms: overview of the development process and project status." in *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2001, p. 662.

[34] L. M. Schriml, C. Arze, S. Nadendla, Y.-W. W. Chang, M. Mazaitis, V. Felix, G. Feng, and W. A. Kibbe, "Disease ontology: a backbone for disease semantic integration," *Nucleic acids research*, vol. 40, no. D1, pp. D940–D946, 2012.

[35] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *arXiv preprint arXiv:1503.00075*, 2015.

APPENDIX

TABLE III

RESULTS (QUADRATIC). MULTIVARIABLE POISSON REGRESSION MODEL WITH FEATURES SORTED BY THE STRENGTH OF THE COEFFICIENT, β. SE: STANDARD ERROR. CI: CONFIDENCE INTERVAL. OASIS: OXFORD ACUTE SEVERITY OF ILLNESS SCORE. SOFA: SEQUENTIAL ORGAN FAILURE ASSESSMENT. ETHN.: ETHNICITY. *RELATIVE TO WHITE

	β	**Odds Ratio** (ϵ^β)	**CI (95%)**	**p-value**
CONTINUOUS FEATURES				
Age (per year)	-2.1E-3	0.99	0.9896 - 0.9904	<0.001
OASIS	2.2E-3	1.002	1.0012 - 1.0028	<0.001
Elixhauser	1.8E-3	0.99	0.9863 - 0.9937	0.92
SOFA	1.8E-2	1.02	1.0178 - 1.0222	<0.001
CATEGORICAL FEATURES				
HIV infection	-6.5E-2	0.94	0.8890 - 0.9910	0.02
Diabetes	-3.4E-2	0.97	0.9547 - 0.9853	<0.001
Ethn. Other*	-1.1E-2	0.99	0.9728 - 1.0072	0.37
Ethn. Hispanic*	-8.7E-3	0.99	0.9547 - 10253	0.51
Ethn. Black*	-4.2E-3	0.99	0.9651 - 1.0149	0.46
Obesity	-6.1E-3	0.99	0.9626 - 1.0174	0.68
Gender (Male)	1.7E-2	1.02	1.0069 - 1.0331	0.004
Cancer	2.1E-2	1.02	0.9947 - 1.0453	0.27
ICU Type (Surgical)	7.3E-2	1.08	1.0637 - 1.0963	<0001
SENTIMENT FEATURES				
Positive	0.37	1.45	1.4486 - 1.4514	<0.001
Negative	0.045	1.05	1.0304 - 1.0696	<0.001
(Positive)^2 (unit^2)	-0.045	0.95	0.9441 - 0.9559	<0.001
(Negative)^2 (unit^2)	-0.002	0.99	0.9892 - 0.9908	<0.001

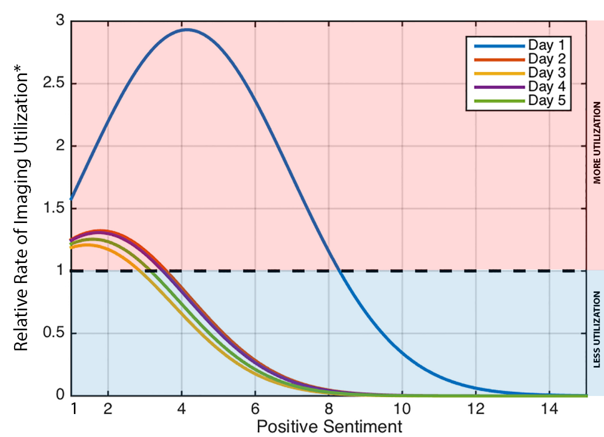

Fig. 5. **Imaging Utilization as a Function of Positive Sentiment.** The effect of positive sentiment (y-axis) on imaging utilization over time for the model with linear and quadratic sentiment terms. Curves represent the estimated relative rate of image utilization per unit increase in positive sentiment, estimated separately for each ICU day. A value of one is interpreted as no effect. The model was trained using all features shown in Appendix Table III in addition to separate day effects, as well as both day **x** sentiment and day **x** sentiment^2 interactions.