

# Using Natural Language Processing to Classify Serious Illness Communication with Oncology Patients

Anahita Davoudi, PhD, MS, MS<sup>1</sup>, Hegler Tissot, PhD<sup>7,8</sup>,  
Abigail Doucette MPH<sup>2</sup>, Peter E. Gabriel MD<sup>2,3</sup>, Ravi Parikh MD, MPP<sup>4</sup>,  
Danielle L. Mowery, PhD, MS, MS<sup>1,2,5\*</sup>, Stephen P. Miranda, MD<sup>6\*</sup>  
<sup>1</sup>Biostatistics, Epidemiology, & Informatics, <sup>2</sup>Abramson Cancer Center,  
<sup>3</sup>Radiation Oncology, <sup>4</sup>Medical Ethics and Health Policy and Medicine,  
<sup>5</sup>Informatics for Biomedical Informatics, <sup>6</sup>Neurosurgery, <sup>7</sup>Computing & Information  
Science; University of Pennsylvania; <sup>8</sup>Information Science, Drexel University

## Abstract

*One core measure of healthcare quality set forth by the Institute of Medicine is whether care decisions match patient goals. High-quality “serious illness communication” about patient goals and prognosis is required to support patient-centered decision-making, however current methods are not sensitive enough to measure the quality of this communication or determine whether care delivered matches patient priorities. Natural language processing (NLP) offers an efficient method for identification and evaluation of documented serious illness communication, which could serve as the basis for future quality metrics in oncology and other forms of serious illness. In this study, we trained NLP algorithms to identify and characterize serious illness communication with oncology patients.*

## 1 Introduction

For patients with cancer to receive care that aligns with their values, their clinicians must effectively explore their care preferences. Documentation of patient-specific goals and prognostic information earlier in the illness trajectory is critical for assessment of shared decision-making, goal-concordance, and healthcare utilization in oncology. High-quality serious illness communication (SIC) can enhance quality of life and goal-concordant care,<sup>1,2,3</sup> while inadequate SIC is associated with greater psychosocial distress and aggressive end-of-life care that may be incongruent with patient preferences.<sup>4,5,6</sup> There is consensus that SIC documentation itself is a core quality measure that supports goal-concordance and therefore must be evaluated.<sup>7,8,9</sup> However, it is well-documented that traditional forms of SIC documentation, including advance directives, are under-utilized and inconsistently applied, making it difficult to track SIC across inpatient and outpatient settings.<sup>10,11,12</sup> High-quality SIC in oncology includes discussion of patient goals, prognosis, code status, and advance care planning.<sup>13</sup> Routine assessment of documentation on these four topics is difficult because this information often exists as free-text in the electronic health record (EHR), which requires time-intensive, manual chart review to identify and abstract.

### 1.1 Natural Language Processing

Natural language processing (NLP) can offer an efficient, accurate alternative for identification of SIC in the EHR<sup>14,15</sup>, and has been used to identify care-planning discussions and palliative care delivery.<sup>16,17,18</sup> Despite early progress, more sophisticated approaches are needed to classify and evaluate SIC documentation. At this time, NLP approaches for identification of SIC predominantly rely on keywords derived from chart review. Such lexical approaches lend themselves well to identification of specific care-planning metrics, such as documentation of code status (e.g. “full code”, “do not resuscitate”) and discussions about hospice (e.g. “comfort measures only”). However, these algorithms are limited in their ability to capture nuanced documentation about patient priorities and prognostic communication, which does not always rely on representative keywords, is less prevalent in the EHR, and is highly variable from clinician to clinician, limiting identification of this documentation at scale.

Machine learning approaches that expand beyond keywords may support more accurate and automatic identification of these two critical SIC domains. In this study, we sought to leverage weakly-labeled EHR data from oncology

---

\*shared co-senior authorship

patients to develop and validate an NLP algorithm that automatically identifies and classifies SIC documentation about prognosis and goals.

## 2 Methods

This study was approved by the University of Pennsylvania Institutional Review Board, protocol #842930. We first collected a weakly annotated dataset of free-text entries containing SIC documentation, and then trained several machine learning algorithms to automatically classify SIC documentation by domain and subdomain. Finally, we characterized the features associated with each SIC subdomain.

### 2.1 Dataset and Schema

In 2018, the University of Pennsylvania Abramson Cancer Center implemented the Serious Illness Care Program (SICP) developed by Ariadne Labs, a multi-component, systems-based intervention designed to enhance timing, frequency, and quality of SIC in oncology.<sup>19,20</sup> Oncology clinicians are encouraged to document SIC using an EHR module, which generates a semi-structured “Serious Illness Conversation” note with subheadings by SIC domain. Prior to this implementation, all clinicians at Abramson Cancer Center were instructed to use an “Advanced Care

**Table 1:** Serious Illness Communication Subdomains for Prognosis and Goals.

Prognosis Domains			
Subdomain	Prompt	Responses	Comment
<b>Prognostic Understanding (PU)</b>	What is your understanding now of where you are with your illness?	<i>Overestimates prognosis;</i> <i>Accurate understanding of prognosis;</i> <i>Underestimates prognosis;</i> <i>No understanding of prognosis;</i>	“He knows he only has weeks to live.”
<b>Information Preferences (IP)</b>	How much information about what is likely to be ahead with your illness would you like from me?	<i>Patient wants to be fully informed;</i> <i>Patient wants to be informed of big picture, but not details;</i> <i>Patient wants some information, but no “bad news”;</i> <i>Patient prefers information to be shared with ***</i>	“She prefers weekly prognosis updates.”
<b>Prognostic Communication (PC)</b>	Information shared with patient about prognosis	<i>Uncertain prognosis;</i> <i>Possibility of getting sick quickly;</i> <i>Limited time, may be as short as ***;</i> <i>May never get stronger or regain function</i>	“He had questions about prognosis.”

Goal Domains			
Subdomain	Prompt	Responses	Comment
<b>Main Goals (MG)</b>	If your health situation worsens, what are your most important goals?	<i>Live as long as possible;</i> <i>Pursue every available treatment;</i> <i>Avoid hospitalizations/maximize time at home;</i> <i>Not be a burden/maintain independence;</i> <i>Be physically comfortable;</i> <i>Be mentally aware;</i> <i>Spent time with family</i>	“The patient wants to live to see his daughter’s wedding.”
<b>Fears/Worries (FW)</b>	What are your biggest fears and worries about the future with your health?	<i>Pain or other symptoms;</i> <i>Loss of control or dignity;</i> <i>Burdening others;</i> <i>Family concerns;</i> <i>Financial concerns</i>	“He worries about becoming dependent.”
<b>Strengths (ST)</b>	What gives you strength as you think about the future with your illness?	<i>Friends/family;</i> <i>Faith/spirituality;</i> <i>Prior experience with adversity</i>	“Support of family and friends.”
<b>Critical Abilities (CA)</b>	What abilities are so critical to your life that you cannot imagine living without them?	<i>Living independently;</i> <i>Being mentally aware;</i> <i>Interacting with others;</i> <i>Dressing, bathing, toileting;</i> <i>Eating and drinking</i>	“Maintaining ability to interact with others is important.”
<b>Tradeoffs (TO)</b>	If you become sicker, how much are you willing to go through for the possibility of gaining more time?	<i>Anything to prolong life incl. life support &amp; ICU care;</i> <i>Limited hospitalizations, some testing and treatments;</i> <i>No further life-prolonging care</i>	“She doesn’t want to experience any major side effects unless there is a high likelihood of therapeutic benefit.”
<b>Family/Friends (FF)</b>	How much does your family know about your priorities and wishes?	<i>Extensive discussion with family about goals and wishes;</i> <i>Some discussion, but incomplete;</i> <i>No discussion but plans to address these issues;</i> <i>No discussion, wants help talking to family;</i> <i>Does not want family informed</i>	“We talked about how he and his wife might begin to have conversations with their daughters.”

Planning” note template for free-text documentation of SIC. In the new “Serious Illness Conversation” note template, there are nine SIC domains, each with a menu of preset responses to choose from, based on the information acquired from the patient, as well as an optional, free-text comment box to insert free-text that provides more detail. The “Serious Illness Conversation” template outlines nine SIC subdomains, three regarding **prognosis** and six regarding **goals**. The **SIC subdomains** including prompts, the structured *responses* and fictitious, but exemplar free-text statements within the “comments” are listed in **Table 1**.

For this study, we queried the Penn Medicine cancer registry for all patients with stage III or IV cancer who were treated across all Penn-affiliated locations and whose records contained “Serious Illness Conversation” notes within our EPIC Clarity electronic data warehouse. Our cohort consisted of 3563 total patients from which 5,145 notes were identified, containing a total of 8,695 distinct “responses” and “comments”. The dataset was randomly split into 6,964 entries (80%) for training and 1,731 entries (20%) for testing.

## 2.2 Serious Illness Communication Classifier Development and Evaluation

Each entry from our dataset was preprocessed using the spaCy library: removing punctuation, eliminating stopwords, reducing case, and encoding n-grams (n=1-3 words).<sup>†</sup> We also encoded lexical categories using Empath.<sup>21</sup> Empath is an unsupervised tool trained using connotations between words leveraging a neural embedding derived from over 1.8 billion words of modern fiction.<sup>‡</sup> Empath can be utilized to generate lexical categories and contains over 200 built-in, topical and emotional categories generated from common dependency relationships in ConceptNet<sup>22</sup> and Parrot.<sup>23</sup> Topical categories include *money, home, work, religion, health, death, etc.* Emotional categories include *sadness, anger, positive emotion, negative emotion, etc.* Terms within both categories were verified using Amazon Mechanical Turk reviewers.

Using the comments from our training dataset, we trained four machine learning algorithms: Logistic Regression, XGBoost, BERT, and Bio+Clinical BERT.

- **Logistic Regression** learns a logit regression model that explains the relationship between the features and the class. Our model uses exhaustive grid search and L1-regularization to optimize performance while reducing the likelihood of over-fitting due to few training examples, many irrelevant features, and a large number of parameters.
- **XGBoost (extreme gradient boosted trees)** is a gradient descent algorithm that learns to predict the residual errors of prior models while minimizing the loss of adding new models before unifying models to make a final class prediction. These boosting models optimize speed and accuracy while reducing the likelihood of overfitting by penalizing trees and applying proportional shrinking of leaf nodes. The booster parameter was set to gblinear.
- **BERT (bidirectional encoder representations from transformers)** are pretrained deep bidirectional representations from unlabeled text fine-tuned using a “masked language model” that combines both left and right contexts.<sup>24§</sup> We leveraged the pre-trained BERT model to provide the vector representations of the embedding sets<sup>25</sup> which were passed to a drop out layer (drop rate of 0.5); the default parameters were used.
- **Bio+Clinical BERT** is a BERT model that leverages pre-trained language representations initialized from BioBERT, a BERT model generated from PubMed article abstracts and PubMed Central article full texts<sup>26</sup> and then fine-tuned using a clinical corpus of notes (e.g., discharge summaries, physician notes, nursing notes, radiology reports, etc.) from the Medical Information Mart for Intensive Care (MIMIC version III) dataset.<sup>27</sup> The default parameters were used.

Using a data-driven approach, we trained each of the four algorithms as a SIC classifier to classify each comment according to SIC domains of goals or prognosis. As a proof-of-concept, we also trained only the logistic regression algorithm to classify 1 out of 9 possible SIC subdomains.

<sup>†</sup><https://spacy.io/universe>

<sup>‡</sup><https://github.com/Ejhfast/empath-client>

<sup>§</sup><https://github.com/google-research/bert>

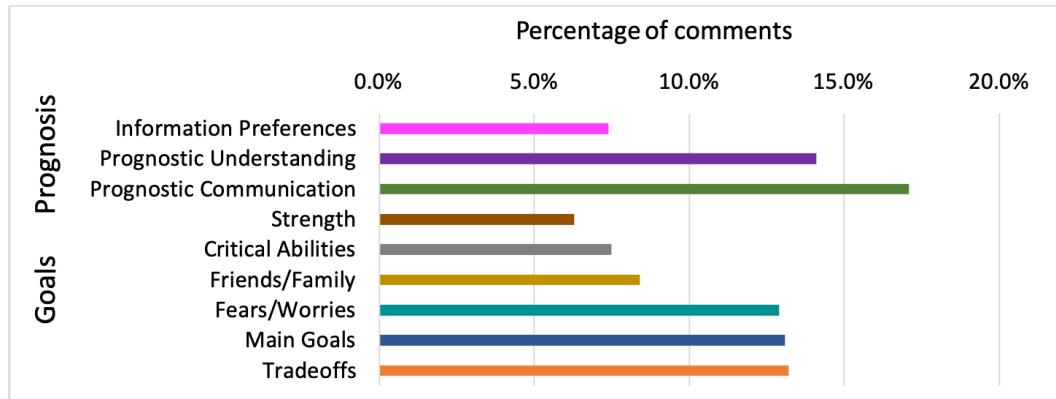
### 2.3 Serious Illness Communication Subdomain Characterization

For each SIC subdomain (e.g., the **Goals** domain has a subdomain of **Strengths**), we applied chi-square feature selection and selected the most significantly associated features (n-grams and Empath categories with  $p < 0.05$ ) associated to each class and applied a log-10 transform to each feature's p-value. We visualized the associated features by transformed p-value using WordCloud. We also report and compare the distribution of Empath categories across subdomains.

## 3 Results

In this study, we leveraged weakly-labeled EHR data from oncology patients to develop and validate an NLP algorithm that automatically identifies and classifies SIC documentation about prognosis and goals.

In **Figure 1**, we report the percent distribution of comments by subdomain across the full corpus. Among all free-text comments, 61.4% belonged to the domain goals and 38.6% belonged to prognosis. For subdomains within goals, we observed proportions ranging from 6.3% Strengths to 13.2% Tradeoffs. For subdomains within prognosis, we observed proportions ranging from 7.4% Information Preferences to 17.1% Prognostic Communication.



**Figure 1:** Overall distribution of SIC domains and subdomains.

### 3.1 Serious Illness Communication Classifier Development and Evaluation

In **Table 2**, we report the predictive performance of each machine learning algorithm on the test set. The highest F1-score was achieved by XGBoost for both prognosis (0.86) and goals (0.91). XGBoost achieved the highest precision for prognosis (0.86) and highest recall for goals (0.92). Conversely, Bio+Clinical BERT achieved the highest recall for prognosis (0.86) and highest precision for goals (0.92). In terms of deep learning algorithms, for both prognosis and goals, we observed higher recall (+6 points, +8 points) and precision (+16 points, +2 points) using Bio+Clinical BERT over BERT, respectively.

**Table 2:** SIC classifier performance by SIC domain on the test set.

Prognosis	Recall	Precision	F1-score
Logistic Regression (baseline)	0.81	0.85	0.83
XGBoost	0.85	0.86	0.86
BERT	0.80	0.64	0.71
Bio+Clinical BERT	0.86	0.80	0.83
Goals	Recall	Precision	F1-score
Logistic Regression (baseline)	0.91	0.89	0.90
XGBoost	0.92	0.91	0.91
BERT	0.80	0.90	0.84
Bio+Clinical BERT	0.88	0.92	0.90

In **Table 3**, we report the predictive performance of the logistic regression algorithm on the test set for each SIC domain. Among prognosis, the highest F1-score was achieved for Prognostic Understanding (0.61) followed by Prognostic Communication (0.60). Among goals, the highest F1-score was achieved for Critical Abilities (0.71) followed by Strengths (0.65) and Tradeoffs (0.63).

**Table 3:** Logistic Regression SIC classifier performance by SIC subdomain on the test set.

Prognosis	Recall	Precision	F1-score
Prognostic Understanding	0.58	0.64	0.61
Information Preferences	0.44	0.42	0.43
Prognostic Communication	0.57	0.63	0.60
Goals	Recall	Precision	F1-score
Main Goals	0.52	0.68	0.59
Fears/Worries	0.62	0.40	0.49
Strengths	0.75	0.58	0.65
Critical Abilities	0.70	0.71	0.71
Tradeoffs	0.60	0.65	0.63
Friends/Family	0.47	0.27	0.35

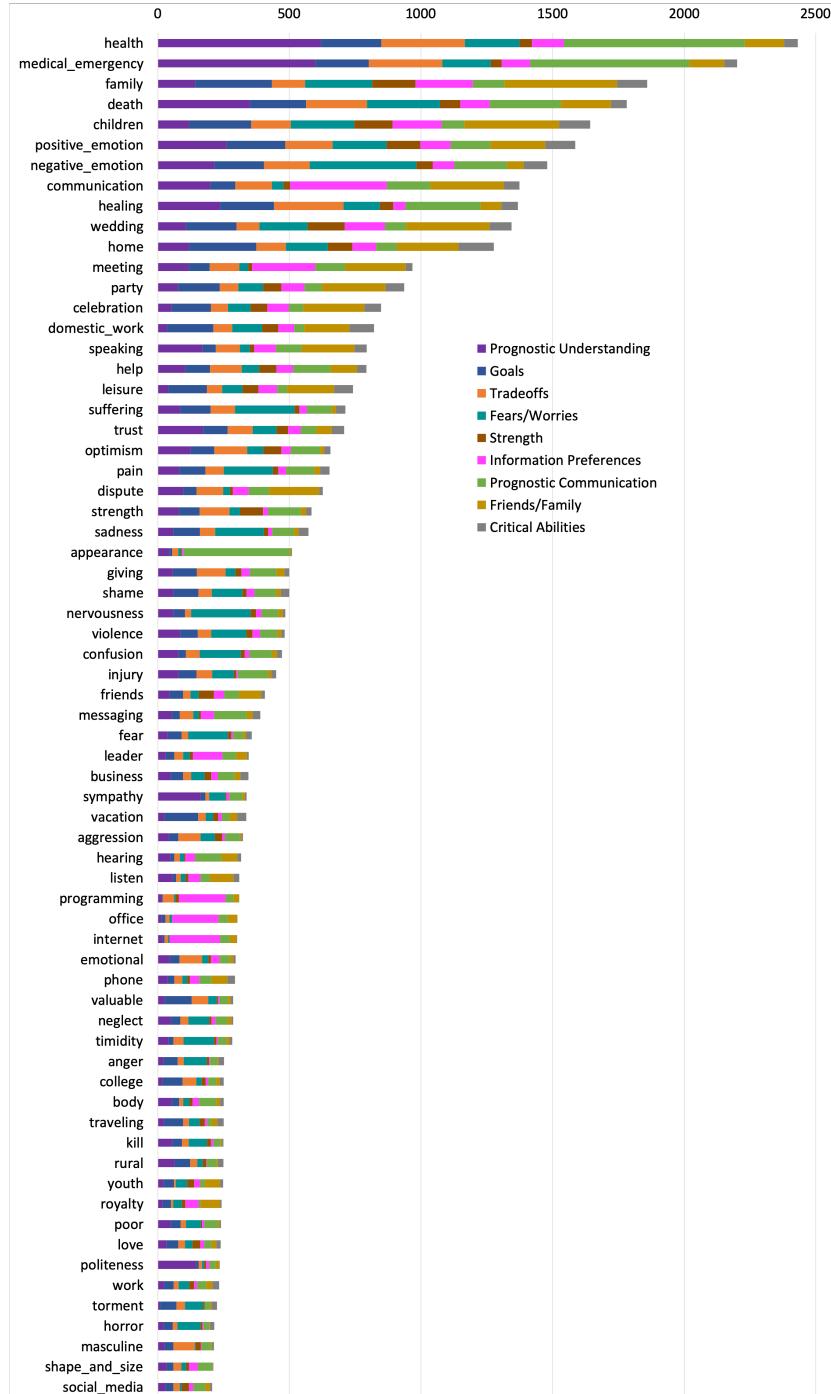
### 3.2 Serious Illness Communication Subdomain Characterization

In **Figure 2**, we present the most informative n-grams and Empath categories associated with each SIC subdomain. Features with high associations (low p-values) to a subdomain are larger in the WordCloud. Notable features by prognosis subdomain include: prognostic understanding (*prognosis, understanding, curable, helpful, understands\_disease, know*), information preferences (*big\_picture, detail, fully\_informed*), and prognostic communication (*limited\_time, short\_months*). Notable features by goal subdomain include: goals (*spend\_time, quality, home, live\_long\_possible, comfortable*), fears/worries (*fear, concern, loss, dying, suffering*), strength (*strength, friends, catholic, spirituality*), critical abilities (*walking, taking\_care, reading, independently, self*), tradeoffs (*intubation, dnr, code, life\_support, would\_want, measures, considering*), and friends/family (*family, extensive, discussion, wife, daughter, conversation*).



**Figure 2:** Wordclouds for Prognosis (1st row) and Goal (2nd/3rd row) Subdomains.

In **Figure 3**, we present the frequency distribution of observed Empath categories according to each SIC subdomain; any Empath category with less than 200 total counts is not shown. We observed 194 of the more than 200 built-in Empath categories in our full dataset. The most common Empath categories observed across the corpus include: *health, medical\_emergency, positive\_emotion, family, children, death, negative\_emotion, and communication*.



**Figure 3:** Frequency distributions of Empath category by subdomain; any Empath category with less than 200 total counts is not shown.

## 4 Discussion

Accurate, reliable, and scalable identification of serious illness communication in the EHR is critical for measuring and improving the quality of oncology care.

### 4.1 Serious Illness Communication Classifier Development and Evaluation

We successfully utilized semi-structured EHR data to develop an NLP algorithm capable of classifying documented entries by SIC domain with high fidelity, identifying text about prognosis (0.86) and goals (0.91). Overall, performance of the classifier across all subdomains ranged from reasonable (0.71) to high (0.91). This study demonstrates promise for identifying SIC—and extracting more complex semantic constructs out of the EHR—without relying on keyword-based approaches. Automated methods for characterizing SIC documentation at scale are limited because clinical notes are variable and often unique to specific clinical situations, which narrow, lexical approaches might fail to anticipate. Here, we leveraged semi-structured data as “weakly labeled” text for classifier training, not only eliminating the need for annotation, but also enhancing the predictive power of the classifier by generating n-grams reflective of diverse lexical categories for training.

The SIC classifier was less effective at discerning individual subdomains within goals and prognosis likely because each subdomain represents overlapping constructs with shared terminology. The “Serious Illness Conversation” template was designed as a communication aid for clinicians to elicit patient values and support prognostic communication, so it is likely that individual subdomains are interrelated for the same patient. While distinguishing between subdomains may be less critical for clinicians using the template at the point of care, enhancing discrimination within each domain would improve classifier performance in free-text clinical notes going forward. It is possible that clinicians inadvertently documented information under the wrong subdomains, which would confound the classifier’s ability to distinguish between them. Notably, classifier performance identified goals better than prognosis, despite a broader range of subdomains, although this may be because the majority of documentation (61.4%) is about goals (**Figure 1**).

The next phase of this research will involve testing and validation of the algorithm’s ability to identify and classify SIC among undifferentiated clinical notes containing unstructured free-text. During this process, further work will be needed to explore why more supervised ML methods (e.g. logistic regression, XGBoost) outperformed deep learning algorithms in this study. Many of the comments and responses used for training and testing consisted of telegraphic phrases, so it may be that deep learning approaches will be more successful in further testing on longer free-text entries, where more contextual features are present. In fact, for both the prognosis and goals domains, we observed higher recall and precision using Bio+Clinical BERT over BERT, respectively, supporting the hypothesis superior performance can be achieved in part through the use of pre-trained models based on clinical documentation.

### 4.2 Serious Illness Communication Subdomain Characterization

Analysis of the most predictive features for each subdomain demonstrates that these features conceptually map very closely to the theme of each subdomain (**Figure 2**) while reflecting a broad range of etymologic categories (**Figure 3**), illustrating the utility of incorporating lexical terms and semantic grouping into the classifier training process. For instance, features associated with documentation about prognosis captured non-specific (*terminal, curable, incurable*) and time-based prognostication (*limited\_time, short\_months, short\_weeks*); the degree of prognostic understanding (*overestimates, accurate, know, good\_understanding, understands\_cancer*); how this information was communicated (*office, internet/email*) and to what extent (*detailed, big\_picture, fully\_informed*).

Similarly, subdomains within goals, features describe specific wishes or priorities (*wedding, quality\_time*) and even place of final rest (*home, die\_house*). Both negative and positive sentiments were reflected. For example, fears/worries contain features of negative emotion (*worried, afraid, suffering, weakness, fearful, nervousness, concern, sadness*); strengths contain features of positive emotion (*comfortable, support, strong*). Sources of strength include one’s faith (*catholic, spirituality, divine*) and support system (*children, friends, family*). Critical abilities highlight activities of leisure (*sports, walking, play, art, driving, reading, working*) and daily living (*living, breathing, eating*) as well as terms related to autonomy (*self, independent, dependence*). To achieve these goals and maintain critical abilities, preferences for life-sustaining treatments were also captured, including code status (*intubation, cpr, dnr, life\_support*,

*resuscitation, ventilation, full\_code).* Both prognosis and goals were often shared with individuals representing family (*wife, husband, son, daughter, sister, son*) and those in decision-making roles (*poa, power of attorney*).

#### 4.3 Clinical Applications

If further validated, the clinical implications of this SIC classifier are compelling. While documentation about goals of care and prognostic communication are known process measures of high-quality palliative care delivery<sup>28</sup>, SIC is poorly captured by administrative claims data, and manual review of individual patient records is laborious and impractical at the population level—yet quality measurement in palliative care is still highly dependent on these two methodologies.<sup>12,29,30,31,32</sup> A validated SIC classifier would offer a powerful tool for more useful quality metrics in oncology, either by evaluating communication quality or developing personalized measures of goal-concordance.<sup>7,33</sup> Reliably tracking patient goals would provide useful context for assessing appropriateness of healthcare utilization, and characterizing narrative arcs in the disease trajectory could help frame quality improvement initiatives and psychosocial interventions during serious illness.<sup>34</sup> In healthcare operations, explainable AI for logistic regression or XGBoost could even be used to inform clinician-facing EHR tools at the point of care, perhaps by visualizing positive coefficients or SHAP values across terms and Empath categories.

Although these results are preliminary, the methodology employed here allows for greater real-world applicability than other reports of NLP approaches to SIC identification thus far, which have all been keyword-based.<sup>15,16,17,18</sup> Recent applications of these methods have seen success in patient groups drawn from pragmatic trials in oncology,<sup>35,36</sup> but due to their lexical basis these efforts have required manual annotation of hundreds of clinical notes, and may be weighted towards inpatient admissions or medical crises requiring treatment decisions.<sup>35</sup> Our method may lay the foundation for more nuanced identification of patient-specific priorities and prognostic communication more upstream in the disease trajectory, which would have significant utility across a wide array of clinical contexts.

#### 4.4 Limitations and Future Work

This study has notable limitations. The SIC classifier was trained using semi-structured Epic EHR modules, which limits the replicability of this work in other settings where source text enriched with SIC may be lacking. Moreover, most SIC documentation in oncology exists within free-text clinical notes, requiring discrimination between relevant and irrelevant text. Performance may suffer in population-level datasets where SIC represents a minority of clinical documentation. In the next phase of this research, classifier training must be enhanced for application to free-text clinical notes. As a first step, we are actively applying the XGBoost classifier for goals and prognosis to sentences from free-text, de-identified clinical neuro-oncology notes that were manually annotated as part of ongoing research and quality improvement efforts at our institution.<sup>37,38</sup> Preliminary results are promising (goals – F1: 0.72; prognosis – F1: 0.70). We anticipated a drop in performance because the schema used for annotation of these notes introduced additional subdomains under goals and prognosis for greater precision.<sup>37,38</sup> Additional training and tuning will be needed to optimize the classifier for free-text notes and additional subdomains, which we plan to complete in the near future by leveraging free-text “Advance Care Planning” (ACP) notes obtained from our EHR.

This classifier is based on documentation from a limited number of oncology clinicians at one institution requiring further study in larger, more diverse populations to assess generalizability. In the future, we aim to better understand how patient preferences evolve over time, as well as any similarities or differences in SIC across gender, race, ethnicity, and culture.<sup>39</sup>

#### Conclusion

Here we describe a novel application of NLP for classifying SIC documentation in oncology. If further validated, such an algorithm can retrieve and evaluate SIC documentation in routine clinical practice as a quality metric<sup>40</sup> to assess key clinical and systems priorities in oncology.

## Acknowledgements

We extend our gratitude to the University of Pennsylvania for partially supporting this important research through Dr. Mowery's start-up funding.

## References

- [1] Mack JW, Weeks JC, Wright AA, Block SD, Prigerson HG. End-of-life discussions, goal attainment, and distress at the end of life: predictors and outcomes of receipt of care consistent with preferences. *Journal of Clinical Oncology*. 2010;28(7):1203.
- [2] Mack JW, Cronin A, Taback N, Huskamp HA, Keating NL, Malin JL, et al. End-of-life care discussions among patients with advanced cancer: a cohort study. *Annals of Internal Medicine*. 2012;156(3):204-10.
- [3] Wright AA, Zhang B, Ray A, Mack JW, Trice E, Balboni T, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. *JAMA*. 2008;300(14):1665-73.
- [4] Detering KM, Hancock AD, Reade MC, Silvester W. The impact of advance care planning on end of life care in elderly patients: randomised controlled trial. *BMJ*. 2010;340.
- [5] Nicholas LH, Langa KM, Iwashyna TJ, Weir DR. Regional variation in the association between advance directives and end-of-life Medicare expenditures. *JAMA*. 2011;306(13):1447-53.
- [6] Teno JM, Gruneir A, Schwartz Z, Nanda A, Wetle T. Association between advance directives and quality of end-of-life care: A national study. *Journal of the American Geriatrics Society*. 2007;55(2):189-94.
- [7] Halpern SD. Goal-concordant care-searching for the holy grail. *The New England Journal of Medicine*. 2019;381(17):1603-6.
- [8] McGinnis JM, Malphrus E, Blumenthal D, et al. Vital signs: core metrics for health and health care progress. 2015.
- [9] Dzau VJ, McClellan MB, McGinnis JM, Burke SP, Coye MJ, Diaz A, et al. Vital directions for health and health care: priorities from a National Academy of Medicine initiative. *JAMA*. 2017;317(14):1461-70.
- [10] Yadav KN, Gabler NB, Cooney E, Kent S, Kim J, Herbst N, et al. Approximately one in three US adults completes any type of advance directive for end-of-life care. *Health Affairs*. 2017;36(7):1244-51.
- [11] Wilson CJ, Newman J, Tapper S, Lai S, Cheng PH, Wu FM, et al. Multiple locations of advance care planning documentation in an electronic health record: are they easy to find? *Journal of Palliative Medicine*. 2013;16(9):1089-94.
- [12] Curtis JR, Sathiratanacheewin S, Starks H, Lee RY, Kross EK, Downey L, et al. Using electronic health records for quality measurement and accountability in care of the seriously ill: opportunities and challenges. *Journal of Palliative Medicine*. 2018;21(S2):S-52.
- [13] Bernacki RE, Block SD. Communication about serious illness care goals: a review and synthesis of best practices. *JAMA Internal Medicine*. 2014;174(12):1994-2003.
- [14] Yim Ww, Yetisgen M, Harris WP, Kwan SW. Natural language processing in oncology: a review. *JAMA Oncology*. 2016;2(6):797-804.
- [15] Lilley EJ, Lindvall C, Lillemoe KD, Tulsky JA, Wiener DC, Cooper Z. Measuring processes of care in palliative surgery: a novel approach using natural language processing. *Annals of Surgery*. 2018;267(5):823-5.
- [16] Lee KC, Udelsman BV, Streid J, Chang DC, Salim A, Livingston DH, et al. Natural language processing accurately measures adherence to best practice guidelines for palliative care in trauma. *Journal of Pain and Symptom Management*. 2020;59(2):225-32.
- [17] Lindvall C, Lilley EJ, Zupanc SN, Chien I, Udelsman BV, Walling A, et al. Natural language processing to assess end-of-life quality indicators in cancer patients receiving palliative surgery. *Journal of Palliative Medicine*. 2019;22(2):183-7.
- [18] Brizzi K, Zupanc SN, Udelsman BV, Tulsky JA, Wright AA, Poort H, et al. Natural language processing to assess palliative care and end-of-life process measures in patients with breast cancer with leptomeningeal disease. *American Journal of Hospice and Palliative Medicine®*. 2020;37(5):371-6.
- [19] Paladino J, Bernacki R, Neville BA, Kavanagh J, Miranda SP, Palmor M, et al. Evaluating an intervention to improve communication between oncology clinicians and patients with life-limiting cancer: a cluster randomized clinical trial of the serious illness care program. *JAMA Oncology*. 2019;5(6):801-9.

- [20] Pasricha V, Gorman D, Laothamatas K, Bhardwaj A, Ganta N, Mikkelsen ME. Use of the serious illness conversation guide to improve communication with surrogates of critically ill patients. A pilot study. *ATS scholar*. 2020;1(2):119-33.
- [21] Fast E, Chen B, Bernstein MS. Empath: Understanding Topic Signals in Large-Scale Text. *CoRR*. 2016;abs/1602.06979. Available from: <http://arxiv.org/abs/1602.06979>.
- [22] Speer R, Chin J, Havasi C. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. *CoRR*. 2016;abs/1612.03975. Available from: <http://arxiv.org/abs/1612.03975>.
- [23] Shaver P, Schwartz J, Kirson D, O'connor C. Emotion knowledge: further exploration of a prototype approach. *Journal of Personality and Social Psychology*. 1987;52(6):1061.
- [24] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding; 2019.
- [25] Turc I, Chang MW, Lee K, Toutanova K. Well-Read Students Learn Better: On the Importance of Pre-training Compact Models. *arXiv preprint arXiv:190808962v2*. 2019.
- [26] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2019 Sep. Available from: <http://dx.doi.org/10.1093/bioinformatics/btz682>.
- [27] Alsentzer E, Murphy J, Boag W, Weng WH, Jindi D, Naumann T, et al. Publicly Available Clinical BERT Embeddings. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop. Minneapolis, Minnesota, USA: Association for Computational Linguistics; 2019. p. 72-8. Available from: <https://aclanthology.org/W19-1909>.
- [28] Forum NQ. NQF-endorsed palliative care and end-of-life care endorsement maintenance standards. 2012.
- [29] Smith G, Bernacki R, Block SD. The role of palliative care in population management and accountable care organizations. *Journal of Palliative Medicine*. 2015;18(6):486-94.
- [30] Kamal AH, Hanson LC, Casarett DJ, Dy SM, Pantilat SZ, Lupu D, et al. The quality imperative for palliative care. *Journal of Pain and Symptom Management*. 2015;49(2):243-53.
- [31] Heyland DK, Dodek P, You JJ, Sinuff T, Hiebert T, Tayler C, et al. Validation of quality indicators for end-of-life communication: results of a multicentre survey. *CMAJ*. 2017;189(30):E980-9.
- [32] Stephens AR, Wiener RS, Ieong MH. Comparison of methods to identify Advance Care Planning in patients with severe chronic obstructive pulmonary disease exacerbation. *Journal of Palliative Medicine*. 2018;21(3):284-9.
- [33] Sanders JJ, Curtis JR, Tulsky JA. Achieving goal-concordant care: a conceptual model and approach to measuring serious illness communication and its impact. *Journal of Palliative Medicine*. 2018;21(S2):S-17.
- [34] Ross L, Danforth CM, Eppstein MJ, Clarfeld LA, Durieux BN, Gramling CJ, et al. Story Arcs in Serious Illness: Natural Language Processing features of Palliative Care Conversations. *Patient Education and Counseling*. 2020;103(4):826-32.
- [35] Lee RY, Brumback LC, Lober WB, Sibley J, Nielsen EL, Treece PD, et al. Identifying Goals of Care Conversations in the Electronic Health Record Using Natural Language Processing and Machine Learning. *Journal of Pain and Symptom Management*. 2021;61(1):136-42.
- [36] Lindvall C, Deng CY, Moseley E, Agaronnik N, El-Jawahri A, Paasche-Orlow MK, et al. Natural Language Processing to Identify Advance Care Planning Documentation in a Multisite Pragmatic Clinical Trial. *Journal of Pain and Symptom Management*. 2021.
- [37] Reed-Guy L, Miranda SP, Alexander TD, Biggiani G, Grady MS, Jones JA, et al. Serious Illness Communication Practices in Glioblastoma: An Institutional Perspective. *Journal of palliative medicine*. 2021;(00):1-9.
- [38] Reed-Guy L, Alexander TD, Biggiani G, Miranda SP, O'Connor N. Serious illness communication practices in glioblastoma care at an academic medical center.. American Society of Clinical Oncology; 2019.
- [39] Cain CL, Surbone A, Elk R, Kagawa-Singer M. Culture and palliative care: preferences, communication, meaning, and mutual decision making. *Journal of Pain and Symptom Management*. 2018;55(5):1408-19.
- [40] Parikh RB, Manz C, Chivers C, Regli SH, Braun J, Draugelis ME, et al. Machine learning approaches to predict 6-month mortality among patients with cancer. *JAMA Network Open*. 2019;2(10):e1915997-7.