



Published in final edited form as:

J Pain Symptom Manage. 2022 January ; 63(1): e29–e36. doi:10.1016/j.jpainsymman.2021.06.025.

Natural Language Processing to Identify Advance Care Planning Documentation in a Multisite Pragmatic Clinical Trial

Charlotta Lindvall, MD, PhD^{1,2,3}, Chih-Ying Deng, MD, MS¹, Edward Moseley, BA¹, Nicole Agaronnik, BS^{1,3}, Areej El-Jawahri, MD^{3,4}, Michael K. Paasche-Orlow, MD^{5,6}, Joshua R. Lakin, MD^{1,2,3}, Angelo Volandes, MD^{3,4,6}, The ACP-PEACE Investigators, James A. Tulsky, MD^{1,2,3}

¹Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute, Boston, MA, USA

²Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

³Harvard Medical School, Boston, MA, USA

⁴Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

⁵Department of Medicine, Boston University School of Medicine, Boston Medical Center, Boston, MA, USA

⁶ACP Decisions, Boston, MA, USA

Abstract

Context: Large multisite clinical trials studying decision-making when facing serious illness require an efficient method for abstraction of advance care planning (ACP) documentation from clinical text documents. However, the current gold standard method of manual chart review is time-consuming and unreliable.

Objectives: To evaluate the ability to use natural language processing (NLP) to identify ACP documentation in clinical notes from patients participating in a multisite trial.

Corresponding Author: Charlotta Lindvall, MD, PhD, Dana-Farber Cancer Institute, 450 Brookline Ave, LW-670, Boston, MA, USA. charlotta_lindvall@dfci.harvard.edu.

Author Contributions:

Study concept and design: Lindvall, Lakin, Volandes, Tulsky

Acquisition of subjects and/or data, analysis and interpretation of data: All authors

Preparation of manuscript: All authors

Collaborators: The ACP-PEACE Investigators: Danielle Kennedy, MPH; Jody-Ann McLeggon, MPH; Camille Chan, BS; Jeremiah J. Stout, BA; Elise N. Brannen, BS, BA; Sophia Zupanc, BA; Daniel A. Gundersen, PhD; Julie Goldman, MS; Susan H. Curtis, MLIS; Haley J. Van Beek, BS; Kathryn I. Pollak, PhD; S. Yousuf Zafar, MD; Jon C. Tilburt, MD; Charles L. Loprinzi, MD; Parvez A. Rahman, MHI; Maria Torroella Carney, MD; Diana Martins-Welch, MD; Michael Qiu, MD, PhD; Edith Burns, MD; Lisa M. Quintiliani, PhD; Aretha Delight Davis, MD, JD

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Data Sharing

For access to ClinicalRegex software contact the corresponding author. Tutorial videos for Clinical Regex can be found on Youtube: <https://www.youtube.com/channel/UC8OcfGj5PkcwG7YU9kIP18g>

Methods: Patients with advanced cancer followed in three disease-focused oncology clinics at Duke Health, Mayo Clinic, and Northwell Health were identified using administrative data. All outpatient and inpatient notes from patients meeting inclusion criteria were extracted from electronic health records (EHRs) between March 2018 and March 2019. NLP text identification software with semi-automated chart review was applied to identify documentation of four ACP domains: (1) conversations about goals of care, (2) limitation of life-sustaining treatment, (3) involvement of palliative care, and (4) discussion of hospice. The performance of NLP was compared to gold standard manual chart review.

Results: 435 unique patients with 79,797 notes were included in the study. In our validation data set, NLP achieved F1 scores ranging from 0.84 to 0.97 across domains compared to gold standard manual chart review. NLP identified ACP documentation in a fraction of the time required by manual chart review of EHRs (1–5 minutes per patient for NLP, vs. 30–120 minutes for manual abstraction).

Conclusion: NLP is more efficient and as accurate as manual chart review for identifying ACP documentation in studies with large patient cohorts.

Introduction

In 2021, more than 1.9 million Americans will receive a new cancer diagnosis and over 600,000 will die from their disease.¹ These patients will accrue hundreds of millions of notes over the course of their cancer care,² including documentation of advance care planning (ACP). The purpose of ACP is to ensure that seriously ill patients receive care consistent with their values and preferences. Further, ACP documentation has been associated with less aggressive care near the time of death.^{3,4} Given the importance of ACP, several ongoing interventions have attempted to increase the quality and frequency of communication about ACP.^{5–8}

Evaluating these ACP interventions, however, requires effective methods for capturing documentation to assess outcomes. Though some ACP documentation can be extracted from electronic health records (EHRs) using medical billing codes and standardized forms, these methods have provided only limited insight. The utility of these methods has been constrained by underutilization structured EHR data elements and over-simplification of complex ACP conversations.^{9–11} Further, these conversations are inconsistently reimbursed by payers and used infrequently by providers.¹² Similarly, advance directives, formal legal documents describing preferences for particular types of medical treatments for serious illness that may be uploaded to the EHR, fall short of painting a robust picture of patients' goals and values.¹³

Given that 70–80% of clinical information in EHRs is text-based,¹⁴ including documentation of in-depth ACP conversations, manual chart review has traditionally served as the only way to extract the data. However, manual extraction is time-consuming, resource-intensive, and requires considerable clinical expertise and training to achieve high inter-rater reliability.¹⁵ Furthermore, abstracting ACP documentation using manual chart review may be infeasible in studies with large sample sizes, such as clinical trials, warranting a more efficient method.

Natural language processing (NLP) presents an opportunity to identify text-based ACP documentation within EHRs with increased speed and efficiency compared to manual chart review.^{16–18} Previous studies have demonstrated that NLP is highly sensitive and specific for identifying serious illness conversations and can reduce the time for data abstraction without compromising quality.^{19–23} The purpose of this paper is to describe a validation study to examine the sensitivity and accuracy of NLP in capturing ACP documentation for a large sample size of patients participating in a multisite pragmatic clinical trial, with the goal of developing a scalable method that can be transferrable to other large pragmatic clinical trials and quality improvement initiatives in the future.

Methods

Sample and Data Sources

Patients included in the study were seen in oncology clinics that participated in the pilot phase (between March 2018 and March 2019) of the pragmatic trial *Advance Care Planning: Promoting Effective and Aligned Communication in the Elderly (ACP-PEACE)* (UG3AG060626-01).⁷ This study spanned three major U.S. healthcare systems: Duke Health, Mayo Clinic, and Northwell Health. All patients were age 65 years or older at the time of their clinic visit and had advanced cancer identified by International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code “secondary malignant neoplasm” C79*.^{24,25} All outpatient and inpatient text documents including history and physical, consult, progress, and discharge notes were extracted from the EHR for each patient for the study period. Duke Health and Mayo Clinic used Epic® in both the inpatient and outpatient setting. Northwell Health used Allscripts™ in the outpatient setting and Sunrise™ in the inpatient setting. This study was approved by the Dana Farber/Harvard Cancer Center Institutional Review Board (18–276).

Natural Language Processing

We used the text annotation software, ClinicalRegex,²⁶ to identify ACP documentation. ClinicalRegex was developed by the Lindvall Lab at Dana-Farber Cancer Institute, and has been applied in multiple studies to assess process-based quality measures.^{21–23,27} Using a pre-defined ontology, the software displays clinical notes that contain highlighted keywords or phrases associated with one or more outcomes of interest. Six human experts (DK, JM, CC, JS, SC, HV) received annotation training through multiple meetings with the senior investigator (CL) to discuss criteria and review coding. Human experts then reviewed clinical notes with highlighted keywords to determine if the context was relevant to the outcome of interest. If the keywords appeared out of context, the notes were labeled for exclusion. This NLP approach allows for semi-automated chart review and reduces the complexity and time required to extract text-based information from EHRs.

Our ontology contained four domains for ACP documentation: 1) goals of care conversation, 2) limitation of life-sustaining treatment, 3) palliative care involvement, or 4) hospice discussion. Using a pre-defined keyword library for ACP documentation developed and validated at Mass General Brigham,²³ the three health care systems and research team participating in our study used an iterative process to further build upon search terms to

reflect institutional-specific language for ACP documentation. The keyword library was refined and validated by expert review of a random selection of notes flagged by NLP, as well as manual review of notes not flagged by NLP. The final keyword library is provided in Table 1. For domains with a low precision metric from at least one health care system, we manually reviewed clinical notes that were identified by NLP but not by human coders.

Performance Metrics

We compared NLP performance to gold standard manual chart review. To construct the gold standard data set for manual chart review, each site manually abstracted a random, representative sample of ACP documentation in the EHRs of a random sample of their patients. These patients represented a subset of the total study population. All clinical notes associated with these patients in the study time frame were abstracted for documentation relevant to each ACP domain. The ACP outcome was met if one or more clinical notes in each patient's EHR contained documentation for at least one ACP domain during the study time period. We used standard metrics to evaluate performance including recall (sensitivity), precision (positive predictive value), accuracy, and F1 score, which assesses the harmonic value between precision and recall. All summary statistics were calculated using Python 3.7.0.

Results

The study population included 435 unique patients with 79,797 notes. The average age was 72 years, 36% were female, and 82% were non-Hispanic white (Table 2).

Of the 435 unique patients in the study population, 60 patients (20 from each site) were randomly selected for construction of the gold standard data set. We used conventional content analysis,²⁸ a qualitative descriptive method,^{29,30} to review EHRs of patients from each site and ensure that a variety of relevant contexts in which keywords may appear are captured in our gold standard data set. Given that we had three disparate healthcare systems in our study, we selected 20 patients from each site to account for potential variations in documentation styles. These 60 patients accrued 8,342 clinical notes in the study time frame, averaging 182 notes per patient (ranging 12 – 267 notes). Given the range of notes associated with each patient, manual abstraction of ACP outcomes from the EHR took 30–120 minutes per patient. NLP analysis of each patient's compiled clinical notes took 1 to 5 minutes. In the gold standard data set, semi-automated chart review with NLP achieved nearly the same performance as manual chart abstraction (F1 score ranging from 0.84 to 1.00 across domains, Table 3).

A review of ACP documentation captured by NLP but not identified in the manual chart review (false positives) at one of our participating healthcare systems with a low precision score demonstrated that additional ACP documentation had been missed by human experts in their manual coding. For example, human experts conducting chart review sometimes missed code status documentation in the limitation of life-sustaining treatment domain, likely because code status was not always reported in the standard location (e.g., at the end of a note). Some notes reported code status in non-standard locations (e.g., only in the History of Present Illness). This occurred with abstraction of code status documentation in

seven clinical notes, where documentation was found using NLP but not during the gold standard manual chart review.

Once the NLP model was validated, the method was applied to the total sample of 435 patients. Of the 435 patients, semi-automated chart review with NLP identified 238 (55%) patients with documentation in at least one domain. For patients with ACP documentation, Figure 1 demonstrates the number of patients with documentation in each domain, patients with documentation in more than one domain, and patients without documentation in any domain. Examples of sample phrases for each domain identified using NLP, as well as examples where keywords appear out of context, are provided in Table 4.

Discussion

Our study demonstrated that NLP can be applied to EHR data to identify ACP documentation with similar performance to gold standard manual chart review. Our NLP method is faster and more accurate than manual review, and may identify documentation that human coders miss. Despite differences in EHRs across healthcare systems, we demonstrated that our NLP method may be scalable for multisite clinical trials. Applying NLP to EHRs offers the potential to efficiently identify important clinical information in large patient cohorts that has previously been infeasible to abstract using manual chart review alone. The limitations of manual chart review have posed a particular challenge for identifying ACP in large cohorts of patients with advanced cancer, given that such documentation often exists in unstructured free text documents. This is the first study to describe a valid, accurate, and efficient NLP method that can identify ACP documentation from such clinical notes in a large multisite clinical trial. This methodology may have further applications for quality improvement efforts and future research studies involving large patient cohorts.

Prior studies have utilized the presence of advance directives and Medical Orders for Life-Sustaining Treatment (MOLST) forms as a method to capture patients' preferences for serious illness care.^{31,32} However, these documents are limited as they provide only a snapshot of the planning activity that may occur during an illness course. In contrast, NLP can be used in real-time to provide data on patients' preferences for ACP throughout their illness trajectory. Given the evolving nature of patients' care preferences, an automated process for tracking such information can ensure that care teams are alerted to important changes in ACP, facilitating timely intervention.

Our finding that NLP sometimes identified ACP documentation missed by manual chart review suggests that the accuracy of NLP may have actually exceeded that of the "gold standard." This underscores the challenges with relying upon manual chart review for abstracting ACP documentation. Reliable human abstraction requires a resource-intensive process including training of human coders to achieve high intra-rater and inter-rater reliability, understanding of the clinical context for ACP conversations, and compensation for the time required. Even with sufficient training of human coders, abstraction of complex data (e.g., adverse events monitoring) from unstructured free text documents may achieve poor inter-rater reliability given the variation in documentation style.^{33,34} Furthermore,

when reviewing large volumes of clinical notes, coders may become prone to specific errors based on phenomena such as pattern recognition to assist with chart abstraction, often looking in standard locations where specific documentation is most likely to be found. This has been observed in previous studies comparing NLP performance to manual chart review, where human coders sometimes missed information because specified keywords were identified in unexpected places (e.g., keywords for “wheelchair” included in medications list).³⁵ Given that human coders may rely upon heuristics for data abstraction, they may miss information that appears in non-standard contexts or due to fatigue. This has also been observed in other studies using diagnostic images, where machine learning performance surpassed human review.³⁶ As an human assisted automated review process, NLP as implemented in the current study may bypass this limitation by reviewing the entire data set with consistent performance, in a fraction of the time required by manual abstraction.

This study has several limitations. Given that writing styles may vary between clinicians, and institutions may have different patient distributions and formats of their health records, our keyword library may not capture all of the variability associated with different documentation styles. Though this study demonstrated that our NLP keyword library efficiently captured ACP documentation across three different institutions and three different types of EHR software, we cannot ensure that this text-extraction approach will be generalizable to all U.S. healthcare settings. Not all U.S. healthcare settings have EHRs that can be readily converted to plain text formats for analysis using ClinicalRegex. Furthermore, generalizability of this NLP keyword library to other institutions will require adaptation and refinement of the current ontology with evaluation of performance in other healthcare settings. Additionally, data identified by NLP is only as useful as the quality of documentation. NLP will not be able to identify important ACP information if it is not recorded in EHRs (i.e., if ACP is only verbally communicated). A potential future applications of our NLP method may include development of a deep learning algorithm that autonomously analyzes whether clinical notes contain keywords that appear in contexts relevant to the outcome of interest. Another future application of our NLP method could be to capture ACP information in real-time from audio-recordings of physician-patient encounters.

Despite these limitations, we have demonstrated that NLP can be used to assess ACP documentation in large EHR datasets that serve pragmatic trials and other clinical uses. The NLP method we developed allows us to study text-based data on a scale that has not previously been possible in the palliative care context. This method may facilitate large scale quality improvement efforts.

ACKNOWLEDGEMENTS

Conflict of Interest:

Dr. Tulskey is a Founding Director of VitalTalk, a non-profit organization focused on clinician communication skills training, from which he receives no compensation. Dr. Volandes has a financial interest in the non-profit foundation Nous Foundation (d/b/a ACP Decisions, 501c3). The non-profit organization develops advance care planning video decision aids and support tools. Dr. Volandes’ interests were reviewed and are managed by Mass General Brigham in accordance with their conflict of interest policies. None of the other authors have any conflicts of interest to disclose.

Sponsor's Role:

The sponsor had no role in the design, data collection, and conduct of this work and did not participate in interpretation of data or preparation of this manuscript.

Funding:

Research reported in this publication was supported within the National Institutes of Health (NIH) Health Care Systems Research Collaboratory by cooperative agreement UH3AG060626 from the National Institute on Aging. This work also received logistical and technical support from the NIH Collaboratory Coordinating Center through cooperative agreement U24AT009676. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. American Cancer Society. Cancer facts and figures 2021. <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2021.html>. Accessed June 1, 2021.
2. Agrawal R, Prabakaran S. Big data in digital healthcare: lessons learnt and recommendations for general practice. *Heredity (Edinb)*. 2020;124(4):525–534. [PubMed: 32139886]
3. Brinkman-Stoppelenburg A, Rietjens JAC, van der Heide A. The effects of advance care planning on end-of-life care: a systematic review. *Palliat Med*. 2014;28(8):1000–1025. [PubMed: 24651708]
4. Fleuren N, Depla MFIA, Janssen DJA, Huisman M, Hertogh CMPM. Underlying goals of advance care planning (ACP): a qualitative analysis of the literature. *BMC Palliat Care*. 2020;19(1):27. [PubMed: 32143601]
5. Walling AM, Sudore RL, Bell D, et al. Population-based pragmatic trial of advance care planning in primary care in the University of California Health System. *J Palliat Med*. 2019;22(S1):72–81. [PubMed: 31486723]
6. Freytag J, Street RLJ, Barnes DE, et al. Empowering older adults to discuss advance care planning during clinical visits: The PREPARE Randomized Trial. *J Am Geriatr Soc*. 2020;68(6):1210–1217. [PubMed: 32157684]
7. Lakin JR, Brannen EN, Tulsy JA, et al. Advance Care Planning: Promoting Effective and Aligned Communication in the Elderly (ACP-PEACE): the study protocol for a pragmatic stepped-wedge trial of older patients with cancer. *BMJ Open*. 2020;10(7):e040999.
8. Grudzen CR, Brody AA, Chung FR, et al. Primary Palliative Care for Emergency Medicine (PRIMER): Protocol for a pragmatic, cluster-randomised, stepped wedge design to test the effectiveness of primary palliative care education, training and technical support for Emergency Medicine. *BMJ Open*. 2019;9(7):e030099.
9. Hunt LJ, Garrett SB, Dressler G, Sudore R, Ritchie CS, Harrison KL. “Goals of care conversations don’t fit in a box”: Hospice staff experiences and perceptions of advance care planning quality measurement. *J Pain Symptom Manage*. October 2020.
10. Reich AJ, Jin G, Gupta A, et al. Utilization of ACP CPT codes among high-need Medicare beneficiaries in 2017: A brief report. *PLoS One*. 2020;15(2):e0228553. [PubMed: 32023311]
11. Gupta A, Jin G, Reich A, et al. Association of billed advance care planning with end-of-life care intensity for 2017 Medicare decedents. *J Am Geriatr Soc*. 2020;68(9):1947–1953. [PubMed: 32853429]
12. Pelland K, Morphis B, Harris D, Gardner R. Assessment of first-year use of Medicare’s advance care planning billing codes. *JAMA Intern Med*. 2019;179(6):827–829. doi:10.1001/jamainternmed.2018.8107 [PubMed: 30855643]
13. Fagerlin A, Schneider CE. Enough. The failure of the living will. *Hastings Cent Rep*. 2004;34(2):30–42. [PubMed: 15156835]
14. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013;309(13):1351–1352. [PubMed: 23549579]
15. Yim W-W, Yetisgen M, Harris WP, Kwan SW. Natural language processing in oncology: A review. *JAMA Oncol*. 2016;2(6):797–804. [PubMed: 27124593]
16. Sager N, Lyman M, Bucknall C, Nhan N, Tick LJ. Natural language processing and the representation of clinical data. *J Am Med Inform Assoc*. 1994;1(2):142–160. [PubMed: 7719796]

17. Stanfill MH, Williams M, Fenton SH, Jenders RA, Hersh WR. A systematic literature review of automated clinical coding and classification systems. *J Am Med Inform Assoc.* 2010;17(6):646–651. [PubMed: 20962126]
18. Dalianis H Applications of clinical text mining. In: *Clinical Text Mining: Secondary Use of Electronic Patient Records.* Cham: Springer; 2018:109–148.
19. Chan A, Chien I, Moseley E, et al. Deep learning algorithms to identify documentation of serious illness conversations during intensive care unit admissions. *Palliat Med.* 2019;33(2):187–196. [PubMed: 30427267]
20. Lee RY, Brumback LC, Lober WB, et al. Identifying goals of care conversations in the electronic health record using natural language processing and machine learning. *J Pain Symptom Manage.* August 2020.
21. Udelsman BV, Lilley EJ, Qadan M, et al. Deficits in the palliative care process measures in patients with advanced pancreatic cancer undergoing operative and invasive nonoperative palliative procedures. *Ann Surg Oncol.* 2019;26(13):4204–4212. [PubMed: 31463695]
22. Poort H, Zupanc SN, Leiter RE, Wright AA, Lindvall C. Documentation of palliative and end-of-life care process measures among young adults who died of cancer: A natural language processing approach. *J Adolesc Young Adult Oncol.* 2020;9(1):100–104. [PubMed: 31411524]
23. Lindvall C, Lilley EJ, Zupanc SN, et al. Natural language processing to assess end-of life quality indicators in cancer patients receiving palliative surgery. *J Palliat Med.* 2019;22(2):183–187. [PubMed: 30328764]
24. Hassett MJ, Ritzwoller DP, Taback N, et al. Validating billing/encounter codes as indicators of lung, colorectal, breast, and prostate cancer recurrence using 2 large contemporary cohorts. *Med Care.* 2014;52(10):e65–73. [PubMed: 23222531]
25. Anand S, Glaspy J, Roh L, et al. Establishing a denominator for palliative care quality metrics for patients with advanced cancer. *J Palliat Med.* 2020;23(9):1239–1242. [PubMed: 31928372]
26. Lindvall Lab. ClinicalRegex. <https://lindvalllab.dana-farber.org/clinicalregex.html>.
27. Lilley EJ, Lindvall C, Lillemoie KD, Tulsy JA, Wiener DC, Cooper Z. Measuring processes of care in palliative surgery: A novel approach using natural language processing. *Ann Surg.* 2018;267(5):823–825. [PubMed: 29112003]
28. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res.* 2005;15(9):1277–1288. [PubMed: 16204405]
29. Sandelowski M Whatever happened to qualitative description? *Res Nurs Health.* 2000;23(4):334–340. [PubMed: 10940958]
30. Sandelowski M What's in a name? Qualitative description revisited. *Res Nurs Health.* 2010;33(1):77–84. [PubMed: 20014004]
31. Tarzian AJ, Cheevers NB. Maryland's Medical Orders for Life-Sustaining Treatment Form use: Reports of a statewide survey. *J Palliat Med.* 2017;20(9):939–945. [PubMed: 28375815]
32. Yadav KN, Gabler NB, Cooney E, et al. Approximately one In three US adults completes any type Of advance directive for end-of-life care. *Health Aff (Millwood).* 2017;36(7):1244–1251. [PubMed: 28679811]
33. Fairchild AT, Tanksley JP, Tenenbaum JD, Palta M, Hong JC. Interrater reliability in toxicity identification: limitations of current standards. *Int J Radiat Oncol Biol Phys.* 2020;107(5):996–1000. [PubMed: 32371073]
34. Atkinson TM, Li Y, Coffey CW, et al. Reliability of adverse symptom event reporting by clinicians. *Qual Life Res.* 2012;21(7):1159–1164. [PubMed: 21984468]
35. Agaronnik ND, Lindvall C, El-Jawahri A, He W, Iezzoni LI. Challenges of developing a natural language processing method with electronic health records to identify persons with chronic mobility disability. *Arch Phys Med Rehabil.* 2020;101(10):1739–1746. [PubMed: 32446905]
36. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018;18(8):500–510. [PubMed: 29777175]

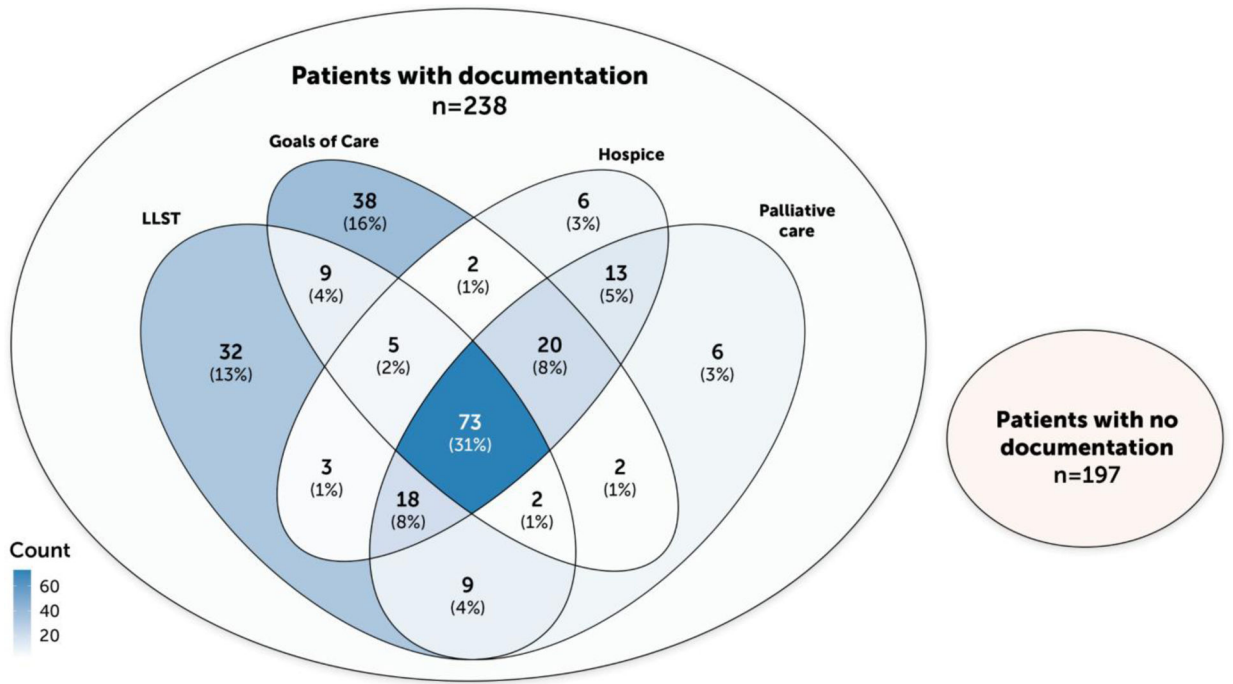


Figure 1:
Venn diagram of patients with documentation in each domain.

Table 1:

Ontology used in NLP text extraction software.

| Domain | Definition | Original keywords from MGB | Keywords added at Mayo | Keywords added at Duke | Keywords added at Northwell |
|---|--|---|--|------------------------|--------------------------------------|
| Goals of care | Documented conversations with patients or family members about the patient's goals, values, or priorities for treatment and outcomes. Includes statements that conversation occurred as well as listing specific goals. OR Documentation that advance care planning was discussed, reviewed, or completed. | goc, goals of care, goals for care, goals of treatment, goals for treatment, treatment goals, family meeting, family discussion, family discussions, patient goals, patient values, quality of life, prognostic discussions, illness understanding, serious illness conversation, serious illness discussion, acp, advance care planning, advanced care planning | supportive care, comfort care, comfort approach, comfort directed care, advanced care plan/goals of care, comfort measures, end of life care | wish | advance care plan, what matters most |
| Limitation of life-sustaining treatment | Documentation about preferences for limitations to cardiopulmonary resuscitation and intubation. | dnr, dnr/dni, dni, dnr/dni, do not resuscitate, do-not-resuscitate, do not intubate, do-not-intubate, no intubation, no mechanical ventilation, no ventilation, no CPR, declines CPR, no cardiopulmonary resuscitation, chest compressions, no defibrillation, no dialysis, no NIPPV, no bipap, no endotracheal intubation, no mechanical intubation, declines dialysis, refuses dialysis, shocks, cmo, comfort measures, comfort, comfort care | do not resuscitate/do not intubate, DNR/DNI/DNH, DNR/I | DNAR | DNR/I |
| Hospice | Documentation that hospice was discussed, prior enrollment in hospice, patient preferences regarding hospice, or assessments the patient did not meet hospice criteria. | hospice | | | |
| Palliative care | Documentation that specialist palliative care was discussed, patient preferences regarding seeing palliative care clinician. | palliative care, palliative medicine, pall care, pallcare, palcare, supportive care | | | |

Abbreviation: MGB, Mass General Brigham

Table 2.

Study sample characteristics stratified by healthcare system.

| Characterics | Overall | Mayo Clinic | Duke Health | Northwell |
|---------------------------|------------|---------------|--------------|------------------|
| Clinic | | Head and Neck | Sarcoma | Gastrointestinal |
| Unique patients, n | 435 | 233 | 74 | 128 |
| Age (years), median (IQR) | 72 (69-77) | 71 (68 – 75) | 73 (70 – 76) | 74 (70–80) |
| Female sex, n (%) | 156 (36%) | 53 (23%) | 34 (46%) | 69 (54%) |
| Non-Hispanic, white n (%) | 355 (82%) | 219 (94%) | 65 (88%) | 71 (56%) |
| Clinical notes, n | 79,797 | 31,981 | 5,718 | 42,098 |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Performance of NLP text extraction method compared to gold standard manual review of EHR.

| | Goals of Care | LLST | Hospice | Palliative Care |
|-------------|---------------|------|---------|-----------------|
| Mayo Clinic | | | | |
| Precision | 1.00 | 1.00 | 1.00 | 0.83 |
| Recall | 1.00 | 1.00 | 1.00 | 1.00 |
| F1 Score | 1.00 | 1.00 | 1.00 | 0.91 |
| Accuracy | 1.00 | 1.00 | 1.00 | 0.90 |
| Duke Health | | | | |
| Precision | 1.00 | 1.00 | 0.67 | 1.00 |
| Recall | 1.00 | 0.80 | 1.00 | 1.00 |
| F1 Score | 1.00 | 0.89 | 0.80 | 1.00 |
| Accuracy | 1.00 | 0.95 | 0.89 | 1.00 |
| Northwell | | | | |
| Precision | 1.00 | 0.33 | 0.50 | 0.83 |
| Recall | 0.88 | 1.00 | 1.00 | 0.83 |
| F1 Score | 0.93 | 0.50 | 0.67 | 0.83 |
| Accuracy | 0.90 | 0.80 | 0.80 | 0.80 |
| Combined | | | | |
| Precision | 1.00 | 0.80 | 0.73 | 0.88 |
| Recall | 0.95 | 0.89 | 1.00 | 0.93 |
| F1 Score | 0.97 | 0.84 | 0.85 | 0.90 |
| Accuracy | 0.97 | 0.92 | 0.89 | 0.92 |

Abbreviation: LLST, Limitation of life-sustaining treatment

Table 4.

Examples of ACP documentation captured by NLP (**Bold** text indicates key words).

| ACP domain | Sample phrase from clinical notes |
|--|--|
| Keywords used in relevant context | |
| Goals of care | Advance Care Planning – Aware disease is incurable, continues to “hope for a cure” if a new therapy arises, but recognizes the reality of her met disease. |
| | GOC/Coping – Processed through with pt his fears that cancer treatment will “take more of a toll” this time around. |
| | We also discussed that we can transition to full comfort care at any time if treatment burdens outweigh potential benefits. |
| | Family meeting held last week patient remains hopeful to return home, improve and be able to undergo another round of chemo. |
| | During goals of care discussion pt was tearful, voicing that she is fearful of what will happen to her husband and their dog. |
| LLST | Code status: DNR /okay to intubate. |
| | Patient is DNR / DNI and goal is to maximize QOL and comfort for as long as possible |
| | In keeping in line with this we discussed changing code status to DNR / DNI . |
| | I am therefore converting his status to DNR / DNI , which is appropriate for this patient with metastatic pancreatic CA. |
| Hospice | Tentative plan for d/c home with hospice later today. |
| | He was quiet but quickly responded not wanting hospice services when his wife brought up the conversation. |
| | Case mgmt to help with hospice screening. |
| Palliative care | Seen in pallcare clinic. |
| | Palliative care following for goals of care discussion with family and consideration of hospice. |
| Keywords used out of context | |
| Goals of care | ... however TPN is dependent on overall GOC as it has not been found to prolong life in patients with terminal illness. |
| | Reason for Consultation: symptom management and goals of care . |
| | ... with consideration of percutaneous nephrostomy if this is within her goals of care . |
| | ... pending GOC discussion – enteral vs. parental if able to place G-J enteral will be preferred. |
| | Also, a family meeting and goals of care conversation is appropriate in the setting of this series of recent admissions for obstructive symptoms. |
| LLST | Pt is feeling at peace with dying and finds a great deal of comfort in her Christian faith. |
| Hospice | Hospice may be suggested depending on course. |
| | We did not specifically talk about hospice care, because they were not yet ready to discuss this, but I would anticipate he would benefit from an referral shortly. |
| Palliative care | Advised supportive care , NGT decompression, observation of LFTs. |
| | Consult: Nutrition, Pall Care |
| | Per oncology clinic note, a pall care consult and social work consult were suggested. |
| | I think that there is no safe next treatment and that we should focus on supportive care . |

Abbreviations: LLST, Limitations of life-sustaining treatment