MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231037

Advanced Care Planning Content Encoding with Natural Language Processing

Benjamin C. KNOLL[a], Melissa GUNDERSON[a], Geetanjali RAJAMANI[a], Elizabeth C. WICK[b], Alexis COLLEY[b], Elizabeth LINDEMANN[a], Rubina RIZVI[a], Molly DIETHELM[a], Gretchen HULTMAN[a], Logan PIERCE[b], Rui ZHANG[a] and Genevieve B. MELTON[a,1]

[a]*University of Minnesota, Minneapolis, MN USA*
[b]*University of California San Francisco, San Francisco, CA USA*
ORCiD ID: Knoll https://orcid.org/0000-0002-1836-721X, Zhang https://orcid.org/0000-0001-8258-3585, Melton https://orcid.org/0000-0001-5193-1663

Abstract. While advanced care planning (ACP) is an essential practice for ensuring patient-centered care, its adoption remains poor and the completeness of its documentation variable. Natural language processing (NLP) approaches hold promise for supporting ACP, including its use for decision support to improve ACP gaps at the point of care. ACP themes were annotated on palliative care notes across four annotators (Fleiss kappa = 0.753) and supervised models trained (Huggingface models *bert-base-uncased* and *Bio_ClinicalBERT*) using 5-fold cross validation (F1=0.8, precision=0.75, recall=0.86, any theme). When applied across the full note corpus of 12,711 notes, we observed variability in documentation of ACP information. Our findings demonstrate the promise of NLP approaches for informatics-based approaches for ACP and patient-centered care.

Keywords. Advanced care planning, natural language processing, patient-centered care, decision support, palliative care

1. Introduction

As the population ages, the number of older adults undergoing medical and surgical care is projected to increase [1]. Illness or major surgery can be associated with complications or worsening health status ranging from temporary functional impairment to unanticipated long-term morbidity and even death, particularly in older adults [2,3]. For example, nearly 20% who die in the U.S. undergo an invasive surgical procedure in the last two months of life [4,5]. Many older adults feel under-prepared for difficult decisions about the intensity and type of treatment that best aligns with their preferences. Some may not have had the chance to rethink their goals in the face of a new major diagnosis. Advanced Care Planning (ACP) discussions can help align treatment intensity with patient preferences to balance risks and benefits of treatments and navigate possible complications. ACP is broadly endorsed as a quality metric, yet real-world ACP adoption remains low particularly in the perioperative setting [6,7].

[1] Corresponding Author: Genevieve B. Melton, MD, PhD, Mayo Mail Code 450, 420 Delaware Street SE, Minneapolis, MN 55455, USA, email:gmelton@umn.edu.

While conversations may be happening between patient and physician, too often these conversations are not documented or are documented in a non-standardized and inaccessible format limiting clinical utility. Furthermore, the heterogeneity of modern health care teams caring for complex patients means that the efficient and timely sharing of clinical information is both increasingly important and complicated. Complex patient-level information, such as overall value placed on functional status versus longevity and important interpersonal relationships, may be contained within free-text notes in an Electronic Health Record (EHR), yet these data remain relatively unusable to coalesce a rich description of patients' preferences with ACP.

Automated methods such as natural language processing (NLP) may help extract data to understand current preferences and goals, changes over time, and to document this in a manner that is clinically useful. While not previously used with ACP elements, NLP methods have been previously used to understand links between utilization of palliative care services and end-of-life quality of life metrics [8] and to measure high-quality palliative care [9,10]. In this study, we sought to extend the applications of NLP to key elements of ACP for information extraction from clinical notes.

2. Methods

To explore the feasibility of developing NLP approaches to extract meaningful data elements from ACP documentation, we applied an established content framework which defines core dimensions of serious illness conversations [11]. Our corpus from M Health Fairview, an academic integrated health system with 10 hospitals based in greater Minnesota and western Wisconsin (United States) (Institutional Review Board approval STUDY00008769) consisted of palliative care notes from 2014-21. Four trained annotators annotated 250 randomly sampled notes with a full overlap of 50 notes to establish inter-rater reliability calculated using Fleiss's kappa.

Table 1. Themes with definitions and example sentences for each theme.

Themes and Definitions	Example sentences
Crit. Abil., Goals, and Tr.. Goals of care, patient treatment priorities, quality-of-life tradeoffs.	She hopes for simple pleasures, normal life, going on drives, eating, enjoying backyard, etc.
Decision-Making. Patient decision-making capability, decision-making preferences.	Defers to his brother [NAME], who in turns connects with extended family members.
Legal Documentation. Health care directive, DNR/DNI, power of attorney, etc.	There is no POLST (Physician orders for life-sustaining treatment) form on file.
Prognosis. Health care personnel prediction of likely course of condition/disease/outcomes.	We talked about how she could live hours to days and possibly even a week.
Understanding. Assessment of health literacy, patient and family understanding of diagnosis, treatment options, and prognosis.	[NAME] does take some time to answer questions, but she has a solid understanding of her health and consequences of her decisions.

Annotation guidelines were made for themes capturing ACP discussions: "Critical Abilities, Goals, and Tradeoffs" ("Crit. Abil., Goals, and Tr."), "Decision-Making", "Legal Documentation", "Prognosis", and "Understanding" (Table 1). Manual labeling was completed at the word level and then mapped to sentence units automatically detected via the BioMedICUS sentence segmentation processor [12]. After annotating 10-note segments of the corpus, annotators reviewed all annotations and resolved any disagreements by consensus. Mapped sentence examples are also illustrated in Table 1.

We used the publicly available Huggingface models *bert-base-uncased* and *Bio_ClinicalBERT* as base models for fine-tuning with our manually annotated corpus as the gold-standard for supervised learning. The *bert-base-uncased* model is a 110M parameter BERT language model trained on the masked language modeling and next sentence prediction tasks, which was trained on BookCorpus, a dataset containing 11,038 unpublished books and English Wikipedia [13]. The *Bio_ClinicalBERT* model is also a 110M parameter BERT language model and was further pre-trained on PubMed data and notes from the MIMIC-III database [14].

We made two modifications to the standard Huggingface Transformers sequence classification task and Trainer class. First, to counteract heavy negative-support class imbalance, positive examples were weighted in inverse proportion to their fraction of support (Table 2). Second, a custom output layer for the multi-label classification problem was used where each theme is optimized against a separate binary cross-entropy loss, which is then summed together to get the overall loss. The model outputs are likelihoods of the sentence containing the themes. Both the target example sentence and a prior sentence are inputs into the model. Training was accomplished using a lightly modified fine-tuning script from the Huggingface project utilizing the transformers Trainer class and unchanged default model parameters. Changes from default training hyperparameters include increasing the batch size from 8 to 32, the gradient accumulation steps from 1 to 16, and the learning rate from 5e-5 to 8e-5. The effect is averaging learning over more training examples, helping to ensure steady convergence despite the low level of class support and presence of multiple learning objectives.

Training and validation were done using 5-fold cross validation on the 250 manually annotated, gold-standard documents. Training was run for 15 epochs taking approximately 1.5 hours per fold utilizing a single nVidia V100. We then examined ACP content across the full corpus to further characterize the ACP theme distribution and completeness within palliative care notes.

3. Results

Between 2014-21, there were 12,711 palliative care notes. Amongst 4 annotators, there was good overall inter-rater reliability across the 50 notes (Table 2, Fleiss's Kappa = 0.753). Across the 250-note evaluation set containing 76,712 sentences, performance for *bert-base-uncased* and *Bio_ClinicalBERT* models (Tables 3 and 4) was similar at the document and sentence level. Across the entire set of notes, we observed variation in the frequency of documentation of ACP themes, namely: "Crit. Abil., Goals, and Tr."; "Decision-Making"; "Legal Documentation"; "Prognosis"; and "Understanding"; or any ACP theme (85.8%, 80.8%, 84.8%, 73.1%, 66.6%, 93.6%, respectively).

Table 2. Fleiss' Kappa interrater agreement (sentence, document) with class weights and support (sentence).

Theme	Sentence-level	Document-level	Class Weight	Class Support
Crit. Abil., Goals, Tr.	0.577	0.650	82.11	923 (1.20%)
Decision-Making	0.747	0.847	175.76	434 (0.57%)
Legal Documentation	0.908	0.849	75.18	1007 (1.31%)
Prognosis	0.586	0.595	225.96	338 (0.44%)
Understanding	0.516	0.492	307.08	249 (0.32%)
Any Theme	0.725	0.794	27.36	2705 (3.53%)

Table 3. Sentence-level mean performance metrics. Cross-validation Std. Dev. in parentheses.

Theme	*bert_base_uncased*			*Bio_ClinicalBERT*		
	Precision	Recall	F1	Precision	Recall	F1
Crit. Abil., Goals, Tr.	.49 (.05)	.75 (.04)	.59 (.03)	.5 (.06)	.76 (.05)	.6 (.05)
Decision-Making	.62 (.07)	.87 (.03)	.72 (.04)	.64 (.07)	.87 (.03)	.73 (.04)
Legal Documentation	.82 (.04)	.95 (.01)	.88 (.02)	.8 (.03)	.96 (.02)	.87 (.03)
Prognosis	.49 (.07)	.73 (.07)	.58 (.06)	.46 (.07)	.73 (.06)	.56 (.07)
Understanding	.39 (.05)	.71 (.08)	.5 (.04)	.38 (.06)	.73 (.1)	.5 (.07)
Any Theme	.75 (.04)	.86 (.02)	.8 (.02)	.75 (.03)	.87 (.03)	.8 (.01)

Table 4. Document-level mean performance metrics. Cross-validation Std. Dev. in parentheses.

Theme	*bert_base_uncased*			*Bio_ClinicalBERT*		
	Precision	Recall	F1	Precision	Recall	F1
Crit. Abil., Goals, Tr.	.81 (.07)	1.0 (.0)	.9 (.04)	.81 (.06)	1.0 (.0)	.89 (.03)
Decision-Making	.9 (.05)	.99 (.01)	.95 (.02)	.9 (.06)	.99 (.01)	.94 (.03)
Legal Documentation	.89 (.07)	1.0 (.0)	.94 (.04)	.88 (.07)	1.0 (.0)	.94 (.04)
Prognosis	.71 (.06)	1.0 (.0)	.83 (.04)	.71 (.08)	1.0 (.0)	.83 (.06)
Understanding	.58 (.07)	.98 (.05)	.73 (.04)	.64 (.05)	.98 (.03)	.78 (.03)
Any Theme	.9 (.04)	1.0 (.0)	.95 (.02)	.9 (.05)	1.0 (.0)	.95 (.03)

4. Discussion

Increasingly, ACP is appreciated as a core resource and part of clinical care through which important, yet difficult, conversations about health care choices occur with patients. Efficient mechanisms to monitor its adoption and to encourage the appropriate use of ACP for decision-making are essential. Similarly, ACP preferences must be communicated to the care team to support patient-centered care. ACP-oriented decision support targeting patients, caregivers, and clinicians is an essential tool for making these improvements. NLP approaches, such as those described in the current evaluation, provide promise and are foundational enablers to informatics approaches for improving the overall adoption and quality of ACP documentation and its use.

While NLP methods performed well, we observed variability in ACP completeness in palliative care notes, particularly for "Prognosis" and "Understanding". Not surprisingly, "Legal Documentation" was the most complete of all of ACP themes. We also observed challenges in getting objective and consistent annotations for certain ACP themes, such as "Understanding" and "Prognosis", which points to some ambiguity to certain ACP themes. Limitations include the use of only two NLP models and the relatively small size of our dataset. Our findings are similar to others [15,16], but our evaluation adds broadly in examining ACP themes in greater detail and describing NLP model performance in detail and is an important foundational step for more robust communication of ACP information. Future work includes external validation and extension of these approaches and leveraging structured data sources. Ultimately, we envision ACP informatics solutions delivering patient-centered care.

5. Conclusions

Our evaluation of NLP and information extraction of ACP from palliative care notes demonstrated good performance and also variability in the completeness of ACP documentation. Further development of these approaches has broad applicability for improved quality of care in key patient populations including older adults.

Acknowledgements

We would like to thank the University of Minnesota Natural Language Processing-Information Extraction Program and the Center for Learning Health System Sciences, as well as funding from NIH (UL1TR000114) and AHRQ (R01HS24532).

References

[1] Hall MJ, Schwartzman A, Zhang J, Liu X. Ambulatory Surgery Data From Hospitals and Ambulatory Surgery Centers: United States, 2010. Natl Health Stat Rep. 2017 Feb;(102):1-15.

[2] Schwarze ML, Barnato AE, Rathouz PJ, Zhao Q, Neuman HB, Winslow ER, Kennedy GD, Hu YY, Dodgion CM, Kwok AC, Greenberg CC. Development of a list of high-risk operations for patients 65 years and older. JAMA Surg. 2015 Apr;150(4):325-31, doi: 10.1001/jamasurg.2014.1819.

[3] Nabozny MJ, Kruser JM, Steffens NM, Pecanac KE, Brasel KJ, Chittenden EH, Cooper Z, McKneally MF, Schwarze ML. Patient-reported limitations to surgical buy-in: a qualitative study of patients facing high-risk surgery. Ann Surg. 2017 Jan;265(1):97-102, doi:10.1097/sla.0000000000001645.

[4] Mohanty S, Rosenthal RA, Russell MM, Neuman MD, Ko CY, Esnaola NF. Optimal perioperative management of the geriatric patient: a best practices guideline from the American College Of Surgeons NSQIP and the American Geriatrics Society. J Am Coll Surg. 2016 May;222(5):930-47, doi:10.1016/j.jamcollsurg.2015.12.026.

[5] Teno JM, Gozalo PL, Bynum JP, Leland NE, Miller SC, Morden NE, Scupp T, Goodman DC, Mor V. Change in end-of-life care for Medicare beneficiaries: site of death, place of care, and health care transitions in 2000, 2005, and 2009. JAMA. 2013 Feb;309(5):470-7, doi:10.1001/jama.2012.207624.

[6] Colley A, Lin JA, Pierce L, Finlayson E, Sudore RL, Wick E. Missed opportunities and health disparities for advance care planning before elective surgery in older adults. JAMA Surg. 2022 Oct;157(10):e223687, doi:10.1001/jamasurg.2022.3687.

[7] Yadav KN, Gabler NB, Cooney E, Kent S, Kim J, Herbst N, Mante A, Halpern SD, Courtright KR. Approximately one in three US adults completes any type of advance directive for end-of-life care. Health Aff (Millwood). 2017 Jul;36(7):1244-51, doi:10.1377/hlthaff.2017.0175.

[8] Brizzi K, Zupanc SN, Udelsman BV, Tulsky JA, Wright AA, Poort H, Lindvall C. Natural language processing to assess palliative care and end-of-life process measures in patients with breast cancer with leptomeningeal disease. Am J Hosp Palliat Care. 2020 May;37(5):371-6, doi:10.1177/1049909119885585.

[9] Lee RY, Brumback LC, Lober WB, Sibley J, Nielsen EL, Treece PD, Kross EK, Loggers ET, Fausto JA, Lindvall C, Engelberg RA, Curtis JR. Identifying goals of care conversations in the electronic health record using natural language processing and machine learning. J Pain Symptom Manage. 2021 Jan;61(1):136-142.e2, doi:10.1016/j.jpainsymman.2020.08.024.

[10] 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. Ann Surg. 2018 May;267(5):823-5, doi:10.1097/sla.0000000000002579.

[11] Bernacki RE, Block SD. Communication about serious illness care goals: a review and synthesis of best practices. JAMA Intern Med. 2014 Dec;174(12):1994-2003, doi:10.1001/jamainternmed.2014.5271.

[12] Knoll BC, Lindemann EA, Albert AL, Melton GB, Pakhomov SVS. Recurrent deep network models for clinical nlp tasks: use case with sentence boundary disambiguation. Stud Health Technol Inform. 2019 Aug;264:198-202, doi:10.3233/SHTI190211.

[13] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv. 2019 May, doi: 10.48550/arXiv.1810.04805.

[14] Alsentzer E, Murphy J, Boag W, Weng W-H, Jindi D, Naumann T, McDermott M. Publicly available clinical BERT embeddings. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop; 2019 Jun; Association for Computational Linguistics; Minneapolis, Minnesota, USA. p. 72-8, doi: 10.18653/v1/W19-1909.

[15] Lee KC, Udelsman BV, Streid J, Chang DC, Salim A, Livingston DH, Lindvall C, Cooper Z. Natural language processing accurately measures adherence to best practice guidelines for palliative care in trauma. J Pain Symptom Manage. 2020 Feb;59(2):225-32.e2, doi:10.1016/j.jpainsymman.2019.09.017.

[16] Lindvall C, Deng CY, Moseley E, Agaronnik N, El-Jawahri A, Paasche-Orlow MK, Lakin JR, Volandes A, Tulsky TAIJA. Natural language processing to identify advance care planning documentation in a multisite pragmatic clinical trial. J Pain Symptom Manage. 2022 Jan;63(1):e29-36, doi:10.1016/j.jpainsymman.2021.06.025.