

Appendix

Table of Contents

CAREGIVER SURVEY SOURCES	2
NATURAL LANGUAGE PROCESSING	2
SOFTWARE SPECIFICATION	2
OUTCOME ONTOLOGY	2
STATISTICAL ANALYSIS	4
SAMPLE SIZE DETERMINATION	4
STATISTICAL ANALYSIS METHODS	4
TABLE 1: GENERAL NLP KEYWORDS	6
REFERENCES	8

Caregiver Survey Sources

Knowledge: We will measure caregiver knowledge of ACP using a 5-item ACP survey consisting of True/False questions that we have validated and used in our prior work.¹⁻³

Confidence: We will ask caregivers how confident they are that their loved one with ADRD will get the type of medical care they want if they become seriously ill. This is a single question with Likert responses ranging from not at all to very confident. Although not psychometrically tested in large studies, it carries enormous face validity and is the closest approximation we can obtain of goal-concordant care prospectively.

Communication Satisfaction: We will use the 10-item communication subscale of the Consumer Assessment of Health Plans (CAHPS®) and ask caregivers to focus on recent ACP communication.

Decisional satisfaction: We will ask the six-item Satisfaction with Decision Scale (Cronbach's alpha = 0.86).⁴ It includes items such as "The decision I made was the best decision possible for my loved one," and "I am satisfied that my decision was consistent with my loved one's values."

Decisional certainty: We will measure decisional conflict, which attempts to measure uncertainty regarding decision making. The Decision Conflict Scale is a well-validated and commonly used tool.⁵

Natural Language Processing

Software Specification

ClinicalRegex⁶, which is a rule-based NLP software with text annotation capacities developed by the Lindvall Lab at Dana-Farber Cancer Institute, will be used to assess goals of care documentation and resuscitation preferences. Clinical Regex has been applied in multiple studies to assess process-based quality measures.⁷⁻²¹ The software presents all keywords and phrases associated with an outcome of interest, and identified using a pre-defined ontology, to a human who reviews the instance. Human reviewers determine whether the keyword or phrase instance appeared in a context that is indicative of goals of care/resuscitation preference documentation. All human reviewers will undergo multiple trainings with study investigators (Moseley, Das, Sciacca, Lindvall).

Outcome Ontology

The ontology for operationalizing the primary outcome of the trial, i.e., goals of care documentation, has six domains that encompass key aspects of goals of care: (1) goals of care conversations; (2) limitations of life-sustaining treatment; (3) palliative care involvement; (4) hospice conversations; (5) election of a surrogate decision maker; (6) time-limited trials. Four of these domains (with the exception of election of a surrogate decision maker and time-limited trials) have been refined and validated in a separate multisite clinical trial.⁹

To identify instances of this documentation associated with the primary and secondary outcomes, clinical experts will work with site teams to develop a list of keywords and phrases for each concept domain that represents, *a priori*, how the domain is believed to be represented within the clinical notes, when spontaneously written in the course of providing care (Table 1). Here, the goal is to capture the language which is most likely to be used when discussing the concepts we are seeking, and then to employ the ClinicalRegex software to search for the keywords and phrases and to point to this documentation in the clinical notes.

Annotator Training and Validation

Annotation guidelines will be developed within the framework of keyword searches to assist annotators in judging instances of document-embedded keywords from the keyword library as representing the primary or secondary outcome concepts. All research staff who will participate in clinical note review (henceforth referred to as ‘operators’) will have to undergo a two-part training process. The first part of the training involves review of software installation, application, and an in-depth review of concepts associated with the study outcome.

Prior to annotating, all operators must earn a passing score on a ‘Calibration Test’ developed by the Lindvall Lab. The ‘Calibration Test’ consists of mock clinical narratives and will test the reviewer’s ability to judge goals of care documentation and resuscitation and treatment preferences. Results of the Calibration Test will be assessed across all concept domains, such that operators may develop proficiency in each.

Operators who completed the Calibration Test (with an accuracy score of 70% or greater) will be considered to have learned the applicable concepts and passed the test. Operators who do not pass the Calibration Test will be re-trained and re-administered the calibration test one more time.

After the calibration tests are completed, at least two operators at each site will annotate a randomly sampled set of notes from 20 unique patients from the study’s baseline period. The operators will annotate these notes three separate times: (1) for concepts associated with the primary outcome; (2) for concepts associated the secondary outcome; (3) for both types of concepts at the same time.

By having operators review notes multiple times, we will generate inter- and intra-operator error rates. Intra-operator error is expected to remain low to ensure that operators are consistent with their own annotation efforts over time. Inter-operator error should also remain low to ensure that annotators are consistent in their understanding of the concepts relative to other operators at the institution. If either intra- or inter operator error are deemed to be too high, re-training on those concepts associated with high rates of error will be performed.

During this validation process, we will also seek to ensure that our keyword library (which is used to search notes for terms and phrases) has as few off-target effects as possible. Specifically, we will seek to ignore keywords and phrases which are common in clinical notes, but are unlikely to be associated with study outcomes. These changes to the keyword libraries will be specific to each site to ensure we account for any site-specific terminology or peculiar documentation practices.

When new operators are onboarded to participate in the annotation effort, those individuals will also participate in the same trainings, and take the calibration test, before annotating the same validation note set. By annotating the same validation note set, we can ensure that new note annotators have similar note annotation practices to their peers.

Statistical Analysis

Sample Size Determination

Data for Aim 1 is derived from clinical notes recorded in the electronic health record (EHR) and as is typical for trials that integrate new initiatives within the workflow of large institutions in a SW-CRT that does not involve consent. We anticipate the sample size for Aim 1 to exceed what is required by a simple application of the power calculation presented above. However, this is warranted for eight reasons. First, the size of this observed sample gives us the opportunity to examine intervention effects for less common outcomes. Second, this sample size will allow us to evaluate potential heterogeneity in treatment effects for subpopulations as small as 20% of the larger study sample. Third, this sample size provides an experimental context in which we will be able to recruit a sample of 500 patients with ADRD or other cognitive impairments and their associated caregivers for survey. In order to sustain the activities of Aim 2, we need a large sample as many people in the sample for Aim 1 will not be eligible for participation in Aim 2. Fourth, the size of the sample for Aim 1 protects this trial from the potential that we will have significantly varying sizes of study clusters, as hospital units vary significantly in their numbers of available patients. This factor is often neglected in sample size assessments for SW-CRTs.²² Fifth, there is minimal risk to human subjects presented by the expanded sample size for Aim 1. Indeed, this educational intervention is being spread across the clinical units of our two hospitals in a pragmatic manner as part of the standard of care.

The research activities of Aim 1 involve no direct burden to patients as there is no consent process and data for this activity will be derived from the EHR. The chief risk is the loss of confidentiality and robust protections are in place to protect patients from this potential risk. Sixth, we plan to extend this intervention as a new clinical initiative in our two health systems in a manner (time per cluster) that has been endorsed by leadership as a reasonable rate for dissemination (i.e., we are not adding more time). Seventh, we have devised an exceedingly efficient and accurate method for outcome assessment (i.e., we are not adding more cost). Eighth, we will protect against inappropriate conclusions. We understand that treatment effect sizes will be more relevant than p-values and that clinical significance is the goal (not simply statistical significance). In summary, the sample size for Aim 1 is needed to be able to address Aim 2 and we have taken appropriate measures to ensure that the research design for Aim 1 does not yield consequences for being overpowered.

Statistical Analysis Methods

Aim 1. To test the combined effects of a PCE-led, video-assisted palliative care intervention on rates of: goals of care documentation; medical orders for resuscitation preferences in the EHR; discussions of palliative care consults; and, discussions of hospice use.

Hypothesis: A higher proportion of patients in the intervention phase (vs. control) will: complete goals of care documentation (primary trial outcome), have documented orders for resuscitation preferences, have documentmation of discussions regarding palliative care consults, and documentmation of discussions regarding hospice.

Based on our prior work in which we exhibited the fact that African-American and Hispanic patients are at particularly high risk for lower level of knowledge related to ACP and goals of care, not discussing goals of care with family, not having a health care proxy, and not having goals of care documentation, we anticipate that this intervention may be particularly beneficial

for African-American and Hispanic patients.²³⁻²⁵ Accordingly, we will evaluate heterogeneous treatment effects by race and ethnicity and anticipate having adequate diversity in our study population to make such assessments. All data regarding Aim 1 will come from the EHR. Our institutions maintain excellent self-report information regarding race and ethnicity.

We will conduct analyses related to potential effect modification as a step in our model validation process and to identify relationships that can be examined more fully in future research. Should interactions not be found to be statistically significant, we will fit a main effects-only model and use it to formally evaluate confounding by applying a change-in-estimates approach, with a 10% change in estimates being an initial screening criterion. Secondary outcomes: Similar procedures will be undertaken to assess intervention effects for the other EHR derived outcomes (documented of resuscitation preferences, palliative care consults, hospice enrollment, and health care proxies in the EHR).

For our primary analysis, we will consider our primary outcome (goals of care documentation) and our secondary outcomes (discussion of resuscitation preferences, palliative care, hospice use, and health care proxies) only for the patient's index hospitalization. However, because we expect some patients to have multiple rehospitalizations during the same step and may also include intervention time (i.e., crossover design), we will perform secondary analyses on all of our primary and secondary outcomes for each patient reviewing all EHR records from the index hospitalization of the patient until their death (or through study period). We will also perform stratified sensitivity analyses of patients who contribute only to control period vs. patients who contribute only to intervention period vs. those that contribute to both control and intervention periods.

We will conduct the above analyses on all patients 65 or over (regardless if they have ADRD) for our Aim 1 primary and secondary outcomes. Aim 1 outcomes will also be analyzed in the 500 patients with ADRD or other cognitive impairments from Aim 2 separately (since some of these patients will likely be younger than 65).

Aim 2. To characterize detailed caregiver-centered outcomes, including knowledge, confidence in future care, communication and decisional satisfaction, and decisional certainty in a subgroup of 500 patients with ADRD or other cognitive impairments and their caregivers admitted to the hospital. Hypothesis: Caregivers in the intervention phase (vs. control) will have higher knowledge, confidence in future care, improved communication and decisional satisfaction, and less decisional conflict.

Table 1: General NLP Keywords

Domain	General Keywords
Goals of Care	GOC, goals of care, goals for care, family meeting, family discussion, patient goals, patient values, quality of life, prognostic discussions, illness understanding, serious illness conversation, serious illness discussion, advance care planning, ACP, end of life, what matters most, poor prognosis, limited prognosis, prognosis, prognostic, terminal, dying, die, death, incurable, not curable, not curative, non curable, non-curable, non curative, non-curative, no cure, isn't a cure, is no cure, not reversible, non-reversible, treatments are palliative, treatment is palliative, palliative treatment, palliation, extend life, extending life, life-extending, life extending, lengthening life, lengthen life, life-lengthening, life lengthening, life limiting, life-limiting, does not wish to know, does not want to know, hours to days, days to weeks, weeks to months, months to years, prognostic understanding, month left, months left, years left, year left, weeks left, week left, unfortunate, regrettably, I am afraid, frank discussion, frank conversation, honest discussion, honest conversation, difficult conversation, difficult discussion, out of options, no remaining options, no more therapy, no further treatment, no further therapy, supportive care, comfort care, comfort approach, CMO, comfort directed care, prioritize comfort, end of life care, comfort measures, limiting invasive procedures, limit invasive procedures, what matters most
Code Status Limitations	Full code, FC, full intubation, full resuscitation, Intubation, resuscitation, CPR, no intubation, no resuscitation, no CPR, declines CPR, do not intubate, do not resuscitate, DNR/DNI, DNR, DNI, declines intubation, declines cardiopulmonary resuscitation, no chest compressions, no compressions, no defibrillation, no mechanical intubation, refuses intubation, refuses CPR, code status discussion, discussed code status, life support, DNAR, do not attempt resuscitation
Hospice	Hospice, bridge to hospice, home hospice, inpatient hospice, hospice house, hospice at home
Palliative Care	Palliative care, palliative medicine, pall care, pal care, pallcare, palcare, PC
Surrogate Decision Maker	Health care agent, health care proxy, HCP, HCP agent, surrogate, surrogate decision maker, decision maker, proxy, health agent, power of attorney for health care, HCPOA, health care power of attorney, health care agent, HCA, guardian guardianship, court appointed guardian, affirmed proxy, POA, POAH
Time Limited Trial	Time-limited trial, time limited trial, limited trial, TLT
Resuscitation	Full code, FC, full intubation, full resuscitation, Intubation, resuscitation, CPR, no intubation, no resuscitation, no CPR, declines CPR, do not intubate, do not resuscitate, DNR/DNI, DNR, DNI, declines intubation, declines cardiopulmonary resuscitation, no chest compressions, no compressions, no defibrillation, no mechanical intubation, refuses intubation, refuses CPR, code status discussion, discussed code status, life support, DNAR, do not attempt resuscitation

Enteral Feeding	Artificial nutrition, feeding tube, supplemental nutrition, nutrition support, PEG, dohoff, G tube, J tube, GJ tube, no artificial feeding, no feeding tube, declines feeding tube, refuses feeding tube, enteral feeding, gastrostomy tube, NG tube, nasogastric tube, OG tube, orogastric
Hemodialysis	Renal replacement therapy, hemodialysis, HD, iHD, CVVH, AVVH, RRT, hemodialysis not within goals, conservative management, medical management without dialysis, no dialysis

References

1. Eneanya ND, Olaniran K, Xu D, et al. Health Literacy Mediates Racial Disparities in Cardiopulmonary Resuscitation Knowledge among Chronic Kidney Disease Patients. *Journal of health care for the poor and underserved*. 2018;29(3):1069-1082. doi:10.1353/hpu.2018.0080
2. El-Jawahri A, Paasche-Orlow MK, Matlock D, et al. Randomized, Controlled Trial of an Advance Care Planning Video Decision Support Tool for Patients With Advanced Heart Failure. *Circulation*. Jul 5 2016;134(1):52-60. doi:10.1161/CIRCULATIONAHA.116.021937
3. El-Jawahri A, Mitchell SL, Paasche-Orlow MK, et al. A Randomized Controlled Trial of a CPR and Intubation Video Decision Support Tool for Hospitalized Patients. *J Gen Intern Med*. Aug 2015;30(8):1071-80. doi:10.1007/s11606-015-3200-2
4. Wills CE, Holmes-Rovner M. Preliminary validation of the Satisfaction With Decision scale with depressed primary care patients. *Health expectations : an international journal of public participation in health care and health policy*. Jun 2003;6(2):149-59.
5. O'Connor AM. Validation of a decisional conflict scale. *Medical decision making : an international journal of the Society for Medical Decision Making*. Jan-Mar 1995;15(1):25-30. doi:10.1177/0272989X9501500105
6. ClinicalRegex. Lindvall Lab. 2021. <https://lindvalllab.dana-farber.org/clinicalregex.html>
7. Johnson PC, Markovitz NH, Gray TF, et al. Association of Social Support With Overall Survival and Healthcare Utilization in Patients With Aggressive Hematologic Malignancies. *J Natl Compr Canc Netw*. Oct 15 2021;1-7. doi:10.6004/jncn.2021.7033
8. Marziliano A, Burns E, Chauhan L, et al. Patient Factors and Hospital Outcomes Associated With Atypical Presentation in Hospitalized Older Adults With COVID-19 During the First Surge of the Pandemic. *J Gerontol A Biol Sci Med Sci*. Jul 19 2021;doi:10.1093/gerona/glab171
9. Lindvall C, Deng CY, Moseley E, et al. Natural Language Processing to Identify Advance Care Planning Documentation in a Multisite Pragmatic Clinical Trial. *J Pain Symptom Manage*. Jul 14 2021;doi:10.1016/j.jpainsymman.2021.06.025
10. Agaronnik N, Lindvall C, El-Jawahri A, He W, Iezzoni L. Use of Natural Language Processing to Assess Frequency of Functional Status Documentation for Patients Newly Diagnosed With Colorectal Cancer. *JAMA Oncol*. Oct 1 2020;6(10):1628-1630. doi:10.1001/jamaoncol.2020.2708
11. 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*. Oct 2020;101(10):1739-1746. doi:10.1016/j.apmr.2020.04.024
12. Brizzi K, Zupanc SN, Udelsman BV, et al. 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*. May 2020;37(5):371-376. doi:10.1177/1049909119885585
13. Udelsman BV, Lee KC, Lilley EJ, Chang DC, Lindvall C, Cooper Z. Variation in Serious Illness Communication among Surgical Patients Receiving Palliative Care. *J Palliat Med*. Mar 2020;23(3):411-414. doi:10.1089/jpm.2019.0268
14. Lee KC, Udelsman BV, Streid J, et al. Natural Language Processing Accurately Measures Adherence to Best Practice Guidelines for Palliative Care in Trauma. *J Pain Symptom Manage*. Feb 2020;59(2):225-232 e2. doi:10.1016/j.jpainsymman.2019.09.017

15. 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*. Dec 2019;26(13):4204-4212. doi:10.1245/s10434-019-07757-2
16. 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*. Feb 2020;9(1):100-104. doi:10.1089/jayao.2019.0040
17. Udelsman B, Lee K, Qadan M, et al. Management of Pneumoperitoneum: Role and Limits of Nonoperative Treatment. *Ann Surg*. Jul 1 2021;274(1):146-154. doi:10.1097/SLA.0000000000003492
18. Udelsman BV, Corey KE, Lindvall C, et al. Risk factors and prevalence of liver disease in review of 2557 routine liver biopsies performed during bariatric surgery. *Surg Obes Relat Dis*. Jun 2019;15(6):843-849. doi:10.1016/j.soard.2019.01.035
19. 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*. Feb 2019;22(2):183-187. doi:10.1089/jpm.2018.0326
20. Udelsman B, Chien I, Ouchi K, Brizzi K, Tulsy JA, Lindvall C. Needle in a Haystack: Natural Language Processing to Identify Serious Illness. *J Palliat Med*. Feb 2019;22(2):179-182. doi:10.1089/jpm.2018.0294
21. Lilley EJ, Lindvall C, Lillemoe KD, Tulsy JA, Wiener DC, Cooper Z. Measuring Processes of Care in Palliative Surgery: A Novel Approach Using Natural Language Processing. *Ann Surg*. May 2018;267(5):823-825. doi:10.1097/SLA.0000000000002579
22. Xu X, Zhu H, Ahn C. Sample size considerations for stratified cluster randomization design with binary outcomes and varying cluster size. *Stat Med*. Aug 15 2019;38(18):3395-3404. doi:10.1002/sim.8175
23. Volandes AE, Ariza M, Abbo ED, Paasche-Orlow M. Overcoming educational barriers for advance care planning in Latinos with video images. *J Palliat Med*. Jun 2008;11(5):700-6. doi:10.1089/jpm.2007.0172
24. Waite KR, Federman AD, McCarthy DM, et al. Literacy and race as risk factors for low rates of advance directives in older adults. *J Am Geriatr Soc*. Mar 2013;61(3):403-6. doi:10.1111/jgs.12134
25. Eneanya ND, Wenger JB, Waite K, et al. Racial Disparities in End-of-Life Communication and Preferences among Chronic Kidney Disease Patients. *Am J Nephrol*. 2016;44(1):46-53. doi:10.1159/000447097