

Supplemental Online Content

Curtis JR, Lee RY, Brumback LC, et al. Intervention to promote communication about goals of care for hospitalized patients with serious illness: a randomized clinical trial. *JAMA*. Published online May 21, 2023. doi:10.1001/jama.2023.8812

eTable. Intervention effect on occurrence of NLP-identified goals-of-care discussions within specific racial groups and within specific ethnic groups

eText. NLP performance by randomization group or by race and ethnicity

eFigure 1. NLP performance over randomization groups within internal validation sample

eFigure 2. NLP performance over minoritized race or ethnicity within internal validation sample

eReference

This supplemental material has been provided by the authors to give readers additional information about their work.

eTable 1. Intervention effect on occurrence of NLP-identified goals-of-care discussions within specific racial groups and within specific ethnic groups^a

Comparison Groups	n	Usual Care, %	Intervention, %	Difference, coefficient	95% CI	<i>p</i>
Race						
Asian	292	29.5	40.6	10.3	-0.9, 21.4	0.070
Black	316	21.6	32.1	10.9	1.2, 20.6	0.028
Native American	45	25.0	23.8	-9.4	-38.8, 19.9	0.519
Pacific Islander	13	22.2	50.0	33.3	-104.9, 171.6	0.593
White	1,768	31.6	34.0	2.4	-2.0, 6.7	0.292
Unknown race	78	43.9	37.8	-4.0	-27.5, 19.5	0.736
Ethnicity						
Hispanic	150	24.7	36.4	13.0	-1.7, 27.6	0.083
Non-Hispanic	2,347	30.6	34.4	3.8	0.0, 7.6	<0.050 ^b
Unknown ethnicity	15	50.0	28.6	-27.3	-131.7, 77.1	0.573

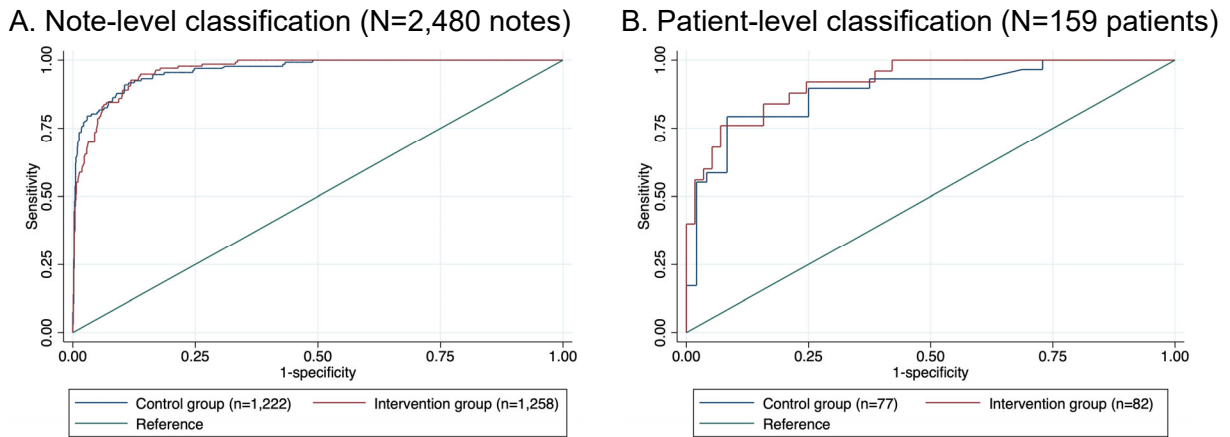
^a The percentages in the two randomization groups are unadjusted. The linear regression results showing the difference (coefficient for randomization condition), *p*, and 95% confidence interval are adjusted for hospital site and for Alzheimer Disease and related dementias (ADRD).

^b *P*-value to 4 decimal places = 0.0497.

eText. NLP performance by randomization group or by race and ethnicity

To examine the possibility of differential NLP performance by randomization group or by race and ethnicity, we conducted *post hoc* receiver operating characteristic (ROC) curve analyses and tested for equality of area under the ROC curve (AUC_{ROC}) over randomization group, and over race and ethnicity using previously described methods.¹ In a random sample of 159 trial participants oversampled for patients with dementia (n=80/159) that underwent complete manual whole-chart abstraction, 1,258/2,480 notes belonged to 82/159 patients in the intervention group, and 1,132/2,470 notes belonged to 73/158 patients of known minoritized race or ethnicity (i.e. non-white race or Hispanic ethnicity). There were no significant differences in NLP AUC_{ROC} at the note or patient levels between randomization groups (eFigure 1) or minoritized race or ethnicity (eFigure 2).

eFigure 1. NLP performance over randomization groups within internal validation sample



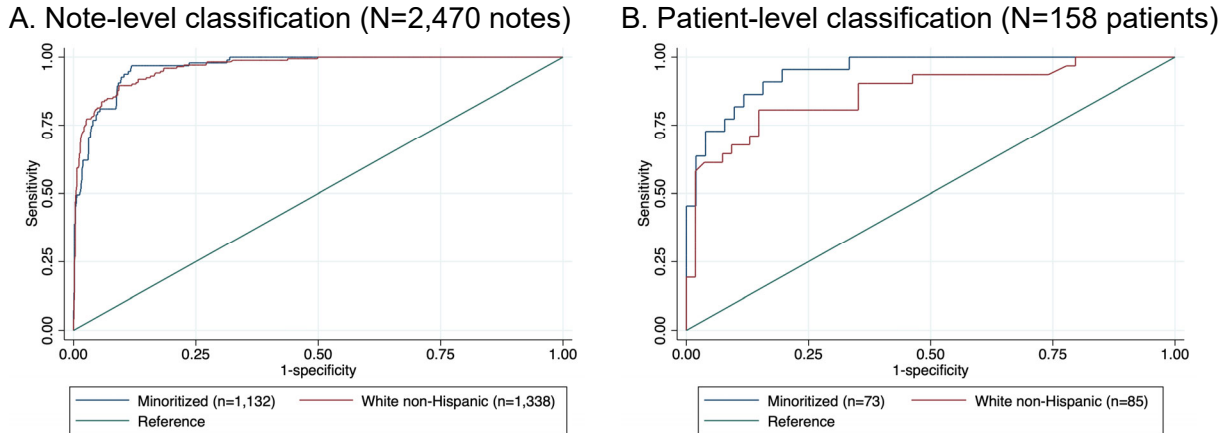
Randomization group	N	n (%) with GOCD ^a	AUC _{ROC} ^b	95%CI	p ^c
A. Note-level classification (N=2,480 notes)					
Control group	1,222 notes	132 (11)	0.964	0.949, 0.979	0.99
Intervention group	1,258 notes	136 (11)	0.964	0.952, 0.975	
B. Patient-level classification (N=159 patients)					
Control group	77 patients	29 (38)	0.887	0.807, 0.967	0.48
Intervention group	82 patients	25 (30)	0.923	0.865, 0.981	

^a Documented goals-of-care discussions (GOCD) as determined by gold standard of manual whole-chart abstraction.

^b Area under the receiver operating characteristic curve for NLP classification of notes or patients with or without documented goals-of-care discussions against human-abstracted gold standard.

^c P-value for test of equality of AUC_{ROC} (H₀: AUC₁=AUC₂) for independent ROC areas.

eFigure 2. NLP performance over minoritized race or ethnicity within internal validation sample



Race and ethnicity ^a	N	n (%) with GOCD ^b	AUC _{ROC} ^c	95%CI	p ^d
A. Note-level classification (N=2,470 notes)					
Non-white or Hispanic	1,132 notes	95 (8)	0.966	0.952, 0.979	0.90
White, non-Hispanic	1,338 notes	172 (13)	0.965	0.952, 0.977	
B. Patient-level classification (N=158 patients)					
Non-white or Hispanic	73 patients	22 (30)	0.948	0.902, 0.995	0.11
White, non-Hispanic	85 patients	31 (36)	0.869	0.784, 0.953	

^a One patient in the internal validation sample had missing data for race and ethnicity.

^b Documented goals-of-care discussions (GOCD) as determined by gold standard of manual whole-chart abstraction.

^c Area under the receiver operating characteristic curve for NLP classification of notes or patients with or without documented goals-of-care discussions against human-abstracted gold standard.

^d P-value for test of equality of AUC_{ROC} (H₀: AUC₁=AUC₂) for independent ROC areas.

eReference

1. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837-45.