Appendix G. Analyses for Heterogeneity of Treatment Effects

Methods

As in the initial analyses of study outcomes, all regression models related to heterogeneity of treatment effects included clustering of patients under clinicians. Because of the small sample sizes of some of the strata of interest, rather than testing a large set of potential confounders for actual confounding and then adjusting those that showed evidence of actual confounding (the practice used in the initial analyses), we automatically adjusted all models for just 3 covariates: patient age, gender, and racial/ethnic minority status. We looked first at associations of the randomization group with the outcomes of interest within strata (the stratum of interest versus all patients not belonging to that stratum) and then used data from the combined strata to test for significant statistical interaction between the randomization group and the stratification variable, basing a conclusion of significance on $P < 0.05$ for the interaction term.

We used the following fields to identify strata of interest:

- *Patients with cancer*: Qualifying diagnosis of metastatic cancer or inoperable lung cancer, or Charlson comorbidity score flag for metastatic tumor, tumor, lymphoma, or leukemia. All patients with none of these indications were in the noncancer group.
- *Patients with lung disease*: Qualifying diagnosis of oxygen-dependent chronic obstructive lung disease, COPD based on FEV, restrictive lung disease, or cystic fibrosis; or Charlson flag for COPD. All patients with none of these indications were in the non–lung disease group.
- *Patients with heart disease*: Qualifying diagnosis of class III or class IV heart failure or pulmonary arterial hypertension, or who had a left ventricular assist device or were aged 65+ with an implantable cardioverter-defibrillator; or Charlson flags of CHF or MI. All patients with none of these indications were in the non–heart disease group.
- *Patients with worse self-assessed general health at the time of study enrollment*: Indication of poor or fair health. Patients reporting good, very good, or excellent health were in the better health group.

Although we did not perform stratified analyses based on patients' ratings of their clinicians' quality of communication, we did test the 4-indicator QOC construct for statistical interaction with the randomization group.

Summary of Findings

These analyses exploring heterogeneity of treatment effects did not produce any consistent evidence of such heterogeneity. Some findings were statistically significant at $p < 0.05$, but

these findings were not consistent and are limited by the large number of comparisons in these analyses. Our ability to identify heterogeneity of treatment effects was limited by our sample size; the study was powered to find a difference overall and not among subgroups. Therefore, we make these analyses available to the interested reader but do not feel we can draw conclusions from them.

Treatment Effects by Group

Patients with cancer versus those without cancer (Table G.1):

Occurrence of discussion about end-of-life preferences at target visit:

Significant treatment effects in the predicted direction in both strata ($p < 0.001$); nonsignificantly stronger treatment effect in the cancer stratum ($p = 0.201$).

Concordance between treatment preference and actual treatment 3 months after the target visit:

In addition to the standard adjustments used for all outcomes, we adjusted this model for treatment preference at 3 months; we included in the analysis only patients who indicated a treatment preference of life extension or comfort care. Nonsignificant treatment effects in the predicted direction in both strata ($p = 0.057$ in cancer stratum, $p = 0.215$ in noncancer stratum); nonsignificantly stronger effect in the cancer stratum ($p = 0.516$).

Depression at 3 months, standard PHQ-8 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of depression at baseline; we omitted from the sample patients for whom a baseline depression score could not be computed. Treatment effect was in the nonpredicted direction and statistically significant in the cancer stratum (b = 1.302; $p = 0.030$); in the predicted direction but not significant in the noncancer stratum (b = –0.182; $p = 0.724$). Although the mean depression at follow-up was higher for the intervention group than for the controls, adjustment for depression at baseline revealed lower depression in the intervention group than in controls. Despite the difference in the direction of the effect in the 2 strata, this difference in patterns was not statistically significant (p for the interaction effect = 0.057).

Anxiety at 3 months, standard GAD-7 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of anxiety at baseline; we omitted from the sample patients for whom a baseline anxiety score could not be computed. The treatment effect in the cancer stratum was statistically significant but in the nonpredicted direction (b = 1.623; $p = 0.021$). In the noncancer stratum the treatment effect was in the predicted direction but nonsignificant (b = −0.985; $p = 0.140$). As with the 3-month depression score, in the noncancer stratum, before adjustment for the baseline level, the intervention group had

higher scores than did controls; however, with adjustment for baseline anxiety, the intervention group's anxiety was lower than that of the control group. The difference in treatment effects in the 2 strata was statistically significant ($p = 0.016$).

Depression at 6 months, standard PHQ-8 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of depression at baseline; we omitted from the sample patients for whom a baseline depression score could not be computed. The treatment effect was in the nonpredicted direction but not statistically significant in both strata ($p = 0.531$ in the cancer stratum and $p = 0.831$ in the noncancer stratum), with nonsignificant ($p = 0.795$) between-strata differences in treatment effect.

Anxiety at 6 months, standard GAD-7 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of anxiety at baseline; we omitted from the sample patients for whom a baseline anxiety score could not be computed. In both strata, after adjustment for the baseline anxiety score, the treatment effect was in the predicted direction but not statistically significant ($p = 0.858$ in the cancer stratum and $p = 0.742$ in the noncancer stratum), with nonsignificant ($p = 0.742$) between-strata differences in treatment effect.

Patients with lung disease versus those without lung disease (Table G.2):

Occurrence of discussion about end-of-life preferences at target visit:

Significant treatment effects in the predicted direction in both strata ($p < 0.001$); nonsignificantly stronger treatment effect in the cancer stratum ($p = 0.760$).

Concordance between treatment preference and actual treatment:

In addition to the standard adjustments used for all outcomes, we adjusted this model for treatment preference at 3 months; we included in the analysis only patients who indicated a treatment preference of life extension or comfort care. Nonsignificant treatment effects in the predicted direction in both strata ($p = 0.136$ in lung-disease stratum, $p = 0.179$ in non–lung disease stratum); nonsignificantly stronger effect in the lung-disease stratum ($p = 0.751$).

Depression at 3 months, standard PHQ-8 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of depression at baseline; we omitted from the sample patients for whom a baseline depression score could not be computed. Treatment effect was in the nonpredicted direction but statistically nonsignificant in both strata ($p = 0.296$ in the lung-disease stratum and $p = 0.949$ in the non–lung disease stratum), with the difference in treatment effects for the 2 strata nonsignificant ($p = 0.292$).

Anxiety at 3 months, standard GAD-7 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of anxiety at baseline; we omitted from the sample patients for whom a baseline anxiety score could not be computed. Treatment effects in both strata were nonsignificant: in the nonpredicted direction in the lung disease stratum (b = 0.850; p = 0.388) and in the predicted direction in the non–lung disease stratum (b = −0.258; p = 0.650). The difference in treatment effects in the 2 strata was not statistically significant (p = 0.384).

Depression at 6 months, standard PHQ-8 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of depression at baseline; we omitted from the sample patients for whom a baseline depression score could not be computed. The treatment effect was nonsignificant in both strata (p = 0.218 in the lung disease stratum and p = 0.063 in the non–lung disease stratum). However, after adjustment for baseline depression, the effect in the lung-disease stratum was in the predicted direction and the effect in the non–lung disease stratum was in the nonpredicted direction; this difference in the pattern of effects was statistically significant (p = 0.022).

Anxiety at 6 months, standard GAD-7 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of anxiety at baseline; we omitted from the sample patients for whom a baseline anxiety score could not be computed. In both strata, after adjustment for the baseline anxiety score, the treatment effect was nonsignificant (p = 0.538 in the lung-disease stratum and p = 0.769 in the non–lung disease stratum), with between-strata differences in treatment effect nonsignificant (p = 0.638).

Patients with heart disease versus those without heart disease (Table G.3):

Occurrence of discussion about end-of-life preferences at target visit:

Significant treatment effects in the predicted direction in both strata (p < 0.001); nonsignificantly weaker treatment effect in the heart disease stratum (p = 0.760).

Concordance between treatment preference and actual treatment:

In addition to the standard adjustments used for all outcomes, we adjusted this model for treatment preference at 3 months; we included in the analysis only patients who indicated a treatment preference of life extension or comfort care. Nonsignificant treatment effects in the predicted direction in both strata (p = 0.172 in heart disease stratum, p = 0.140 in non–heart disease stratum); nonsignificantly weaker effect in the heart disease stratum (p = 0.869).

Depression at 3 months, standard PHQ-8 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of depression at baseline; we omitted from the sample patients for whom a

baseline depression score could not be computed. Treatment effect was in the nonpredicted direction but statistically nonsignificant in both strata (p = 0.497 in the heart disease stratum and p = 0.759 in the non–heart disease stratum), with the difference in treatment effects for the 2 strata nonsignificant (p = 0.915).

Anxiety at 3 months, standard GAD-7 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of anxiety at baseline; we omitted from the sample patients for whom a baseline anxiety score could not be computed. Treatment effects in both strata were nonsignificant: in the nonpredicted direction in the heart disease stratum (b = 0.453; p = 0.517) and in the predicted direction in the non–lung disease stratum (b = −0.235; p = 0.747). The difference in treatment effects in the 2 strata was not statistically significant (p = 0.611).

Depression at 6 months, standard PHQ-8 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of depression at baseline; we omitted from the sample patients for whom a baseline depression score could not be computed. In the heart disease stratum, although the depression score at 6 months was higher in the intervention group than among controls, after adjustment for baseline depression the effect was in the predicted direction but nonsignificant (p = 0.547). In the non–heart disease stratum, the effect was in the nonpredicted direction and just shy of statistical significance (p = 0.052). The difference between strata in the pattern of effects was statistically significant (p = 0.029).

Anxiety at 6 months, standard GAD-7 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of anxiety at baseline; we omitted from the sample patients for whom a baseline anxiety score could not be computed. In both strata, after adjustment for the baseline anxiety score, the treatment effect was nonsignificant in the nonpredicted direction with p = 0.376 in the heart-disease stratum and in the predicted direction with p = 0.128 in the non–heart disease stratum). Despite the difference in the direction of effect in the 2 strata, the between-strata difference was nonsignificant (p = 0.070).

Patients with worse versus better self-assessed health at baseline (groups divided at the sample median: worse = poor/fair; better = good/very good/excellent; Table G.4):

Occurrence of discussion about end-of-life preferences at target visit:

Significant treatment effects in the predicted direction in both strata (p < 0.001); treatment effect slightly but not significantly stronger in patients with better health.

Concordance between treatment preference and actual treatment:

In addition to the standard adjustments used for all outcomes, we adjusted this model for treatment preference at 3 months; we included in the analysis only patients who indicated a treatment preference of life extension or comfort care. Nonsignificant treatment effect in the predicted direction in both strata, with $p = 0.054$ among those with poorer health and $p = 0.341$ among those with better health. Nonsignificant difference in treatment effect for the 2 strata (*P* value for interaction = 0.222).

Depression at 3 months, standard PHQ-8 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of depression at baseline; we omitted from the sample patients for whom a baseline depression score could not be computed. With adjustment for baseline depression, the treatment effect was nonsignificant in both strata (in the nonpredicted direction for those with poorer health, $p = 0.194$; in the predicted direction for those with better health, $p = 0.751$). Difference in treatment effect between the 2 strata was nonsignificant ($p = 0.214$).

Anxiety at 3 months, standard GAD-7 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of anxiety at baseline; we omitted from the sample patients for whom a baseline anxiety score could not be computed. Treatment effects in both strata were nonsignificant: in the nonpredicted direction in the poorer-health stratum and in the predicted direction in the better-health stratum, with the difference in treatment effects in the 2 strata nonsignificant ($p = 0.886$).

Depression at 6 months, standard PHQ-8 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of depression at baseline; we omitted from the sample patients for whom a baseline depression score could not be computed. The treatment effect in the poorer-health stratum was in the predicted direction but not significant ($p = 0.439$); in the better-health stratum the effect was in the nonpredicted direction and significant ($p = 0.029$); the difference in treatment effect between the 2 strata was nonsignificant ($p = 0.080$).

Anxiety at 6 months, standard GAD-7 score:

In addition to the standard adjustments used for all outcomes, we adjusted this model for the level of anxiety at baseline; we omitted from the sample patients for whom a baseline anxiety score could not be computed. In both strata, after adjustment for the baseline anxiety score, the treatment effect was nonsignificant (in the predicted direction with $p = 0.529$ in the poorer-health stratum, and in the nonpredicted direction with $p = 0.979$ in the better-health stratum), with the differences between strata nonsignificant ($p = 0.739$).

Interaction of treatment group with patients' baseline ratings of their clinician's quality of communication (Table G.5):

Stratified analyses could not be performed because QOC was a latent construct. However, when that construct was used as a linear predictor along with the randomization group, the interaction between those 2 predictors, and the automatic adjustment variables, the interaction term was nonsignificant for all outcomes (*P* values ranging from 0.261 to 0.743). We found no evidence that the QOC rating acted as an effect modifier.

Table G.1. Association of Randomization Group With Study Outcomes, Comparing Patients With and Without Cancer[a]

Outcome	Patients With Cancer					Patients Without Cancer					Interaction[d]	
	No. of Patients/ No. of Clinicians[b]	Descriptive by Group[c]		β	*p*	Valid n[b]	Descriptive by Group[c]		β	*p*	β	*p*
		Control	Intervention				Control	Intervention				
Aim 1: Events at Target Visit												
Occurrence of discussion[e]	160/74	0.207	0.706	1.492	**< 0.001**	235/91	0.395	0.767	1.052	**< 0.001**	0.404	0.201
Aim 2: Concordance at 3 Months[f]	109/70	0.538	0.705	0.495	0.057	172/81	0.590	0.697	0.281	0.215	0.201	0.516
Aim 3: Depression and Anxiety												
Standard PHQ-8 score, 3 months[g]	143/76	5.281	6.793	1.302	**0.030**	216/88	4.570	5.451	−0.182	0.724	1.457	0.057
Standard GAD-7 score, 3 months[h]	145/74	3.481	3.737	1.623	**0.021**	221/92	2.630	3.011	−0.985	0.140	2.366	**0.016**
Standard PHQ-8 score, 6 months[g]	113/65	5.091	6.231	0.529	0.531	201/89	4.676	5.779	0.118	0.831	0.247	0.795
Standard GAD-7 score, 6 months[h]	119/67	3.689	3.385	−0.148	0.858	208/89	2.698	3.370	−0.221	0.742	−0.349	0.742

a Results were based on complex regression models with patients clustered under clinicians. All models included automatic adjustment for patient age, gender, and racial/ethnic minority status.

b Number of patients/number of clinician clusters.

c For binary outcomes, the descriptives show the proportion of the group with the outcome. For composite scores, the descriptives represent the mean value of the score at follow-up.

d Test for stratification group as an effect modifier of the association between the randomization group and the outcome. This statistic was based on a model using data from both disease groups, and with the randomization group indicator, the binary disease group indicator, and the product of the 2 indicators as predictors, along with the adjustments used in the stratified models. The coefficient and *P* value are for the product term.

e Binary outcome modeled with probit regression, estimated with weighted least squares estimator with mean and variance adjustment.

f Binary outcome (1 = treatment preference and actual treatment at 3 months were both life extension or comfort care; 0 = treatment preference at 3 months was life extension and actual treatment was comfort care, or the reverse; or patient wasn't sure about preference or actual treatment). In addition to the adjustments for patient gender, age, and racial/ethnic minority status, adjustment was automatically made for treatment preference at 3 months (life extension or comfort care); patients with other values on this adjustment variable were excluded.

g Robust linear regression model, estimated with restricted maximum likelihood. In addition to the adjustments for patient gender, age, and racial/ethnic minority status, automatic adjustment was made for the scale score at baseline.

h Tobit regression model (scale score defined as censored from below), estimated with WLSMV. In addition to the adjustments for patient gender, age, and racial/ethnic minority status, automatic adjustment was made for the scale score at baseline.

Table G.2. Association of Randomization Group With Study Outcomes, Comparing Patients With and Without Lung Disease[a]

Outcome	Patients With Lung Disease						Patients Without Lung Disease					Interaction[d]	
	No. of Patients/ No. of Clinicians[b]	Descriptive by Group[c]		β	*p*	Valid n[b]	Descriptive by Group[c]		β	*p*		β	*p*
		Control	Inter-vention				Control	Inter-vention					
Aim 1: Events at Target Visit													
Occurrence of discussion[e]	111/73	0.370	0.825	1.295	**< 0.001**	284/106	0.293	0.709	1.224	**< 0.001**		0.086	0.760
Aim 2: Concordance at 3 Months[f]	81/58	0.564	0.714	0.457	0.136	200/101	0.569	0.692	0.303	0.179		0.116	0.751
Aim 3: Depression and Anxiety													
Standard PHQ-8 score, 3 months[g]	105/69	5.148	7.526	0.935	0.296	254/104	4.786	5.143	0.028	0.949		0.951	0.292
Standard GAD-7 score, 3 months[h]	103/69	3.597	4.761	0.850	0.388	263/107	2.811	2.546	−0.258	0.650		0.846	0.384
Standard PHQ-8 score, 6 months[g]	94/65	5.163	5.790	−0.961	0.218	220/102	4.705	5.988	1.001	0.063		−2.203	**0.022**
Standard GAD-7 score, 6 months[h]	97/66	3.575	4.226	−0.635	0.538	230/103	2.884	2.982	0.181	0.769		−0.546	0.638

a Results were based on complex regression models with patients clustered under clinicians. All models included automatic adjustment for patient age, gender, and racial/ethnic minority status.
b Number of patients/number of clinician clusters.
c For binary outcomes, the descriptives show the proportion of the group with the outcome. For composite scores, the descriptives represent the mean value of the score at follow-up.
d Test for stratification group as an effect modifier of the association between the randomization group and the outcome. This statistic was based on a model using data from both disease groups, and with the randomization group indicator, the binary disease group indicator, and the product of the 2 indicators as predictors, along with the adjustments used in the stratified models. The coefficient and *P* value are for the product term.
e Binary outcome modeled with probit regression, estimated with weighted least squares estimator with mean and variance adjustment.
f Binary outcome (1 = treatment preference and actual treatment at 3 months were both life extension or comfort care; 0 = treatment preference at 3 months was life extension and actual treatment was comfort care, or the reverse; or patient wasn't sure about preference or actual treatment). In addition to the

adjustments for patient gender, age, and racial/ethnic minority status, adjustment was automatically made for treatment preference at 3 months (life extension or comfort care); patients with other values on this adjustment variable were excluded.

g Robust linear regression model, estimated with restricted maximum likelihood. In addition to the adjustments for patient gender, age, and racial/ethnic minority status, automatic adjustment was made for the scale score at baseline.

h Tobit regression model (scale score defined as censored from below) estimated with WLSMV. In addition to the adjustments for patient gender, age, and racial/ethnic minority status, automatic adjustment was made for the scale score at baseline.

Table G.3. Association of Randomization Group With Study Outcomes, Comparing Patients With and Without Heart Disease[a]

Outcome	Patients With Heart Disease					Patients Without Heart Disease					Interaction[d]	
	No. of Patients/ No. of Clinicians[b]	Descriptive by Group[c]		β	*p*	Valid n_b	Descriptive by Group[c]		β	*p*	β	*p*
		Control	Intervention				Control	Intervention				
Aim 1: Events at Target Visit												
Occurrence of discussion[e]	215/98	.382	.752	1.036	**< 0.001**	180/82	0.248	0.732	1.396	**< 0.001**	−0.339	0.276
Aim 2: Concordance at 3 Months[f]	157/84	0.565	0.693	0.331	0.172	124/75	0.570	0.711	0.400	0.140	−0.057	0.869
Aim 3: Depression and Anxiety												
Standard PHQ-8 score, 3 months[g]	201/93	4.262	5.963	0.340	0.497	158/83	5.501	5.855	0.195	0.759	0.082	0.915
Standard GAD-7 score, 3 months[h]	208/96	2.454	3.118	0.453	0.517	158/82	3.588	3.508	−0.235	0.747	0.538	0.611
Standard PHQ-8 score, 6 months[g]	186/92	4.587	5.421	−0.322	0.547	128/72	5.125	6.907	1.583	0.052	−2.125	**0.029**
Standard GAD-7 score, 6 months[h]	190/93	2.349	3.371	0.666	0.376	137/73	3.902	3.382	−1.108	0.128	1.853	0.070

a Results were based on complex regression models with patients clustered under clinicians. All models included automatic adjustment for patient age, gender, and racial/ethnic minority status.

b Number of patients/number of clinician clusters.

c For binary outcomes, the descriptives show the proportion of the group with the outcome. For composite scores, the descriptives represent the mean value of the score at follow-up.

d Test for stratification group as an effect modifier of the association between the randomization group and the outcome. This statistic was based on a model using data from both disease groups, and with the randomization group indicator, the binary disease group indicator, and the product of the 2 indicators as predictors, along with the adjustments used in the stratified models. The coefficient and *P* value are for the product term.

e Binary outcome modeled with probit regression, estimated with weighted least squares estimator with mean and variance adjustment.
f Binary outcome (1 = treatment preference and actual treatment at 3 months were both life extension or comfort care; 0 = treatment preference at 3 months was life extension and actual treatment was comfort care, or the reverse; or patient wasn't sure about preference or actual treatment). In addition to the adjustments for patient gender, age, and racial/ethnic minority status, adjustment was automatically made for treatment preference at 3 months (life extension or comfort care); patients with other values on this adjustment variable were excluded.
g Robust linear regression model, estimated with restricted maximum likelihood. In addition to the adjustments for patient gender, age, and racial/ethnic minority status, automatic adjustment was made for the scale score at baseline.
h Tobit regression model (scale score defined as censored from below), estimated with WLSMV. In addition to the adjustments for patient gender, age, and racial/ethnic minority status, automatic adjustment was made for the scale score at baseline.

Table G.4. Association of Randomization Group With Study Outcomes, Comparing Patients With Different Baseline Levels of Self-Assessed Health[a]

Outcome	Patients With Poor/Fair Health					Patients With Good/Very Good/Excellent Health					Interaction[d]	
	No. of Patients/ No. of Clinicians[b]	Descriptive by Group[c] Control	Descriptive by Group[c] Intervention	β	*p*	Valid n[b]	Descriptive by Group[c] Control	Descriptive by Group[c] Intervention	β	*p*	β	*p*
Aim 1: Events at Target Visit												
Occurrence of discussion[e]	170/100	0.356	0.747	1.108	**< 0.001**	224/102	0.282	0.750	1.313	**< 0.001**	0.190	0.485
Aim 2: Concordance at 3 Months[f]	116/80	0.552	0.759	0.582	0.054	164/ 93	0.578	0.662	0.207	0.341	−0.372	0.222
Aim 3: Depression and Anxiety												
Standard PHQ-8 score, 3 months[g]	151/92	6.605	7.644	0.930	0.194	207/102	3.771	4.424	−0.151	0.751	−0.999	0.214
Standard GAD-7 score, 3 months[h]	156/97	4.116	4.054	0.189	0.820	209/100	2.252	2.527	−0.050	0.934	−0.143	0.886
Standard PHQ-8 score, 6 months[g]	128/90	7.394	7.347	−0.683	0.439	186/ 99	3.173	4.883	1.075	**0.029**	1.692	0.080
Standard GAD-7 score, 6 months[h]	135/92	4.466	4.622	−0.569	0.529	191/100	2.122	2.393	0.015	0.979	0.344	0.739

a Results were based on complex regression models with patients clustered under clinicians. All models included automatic adjustment for patient age, gender, and racial/ethnic minority status.
b Number of patients/number of clinician clusters.

c For binary outcomes, the descriptives show the proportion of the group with the outcome. For composite scores, the descriptives represent the mean value of the score at follow-up.
d Test for stratification group as an effect modifier of the association between randomization group and the outcome. This statistic was based on a model using data from both health-status groups, and with the randomization group indicator, the binary health-status-group indicator, and the product of the 2 indicators as predictors, along with the adjustments used in the stratified models. The coefficient and *P* value are for the product term.
e Binary outcome modeled with probit regression, estimated with weighted least squares estimator with mean and variance adjustment.
f Binary outcome (1 = treatment preference and actual treatment at 3 months were both life extension or comfort care; 0 = treatment preference at 3 months was life extension and actual treatment was comfort care, or the reverse; or patient wasn't sure about preference or actual treatment). In addition to the adjustments for patient gender, age, and racial/ethnic minority status, adjustment was automatically made for treatment preference at 3 months (life extension or comfort care); patients with other values on this adjustment variable were excluded.
g Robust linear regression model, estimated with restricted maximum likelihood. In addition to the adjustments for patient gender, age, and racial/ethnic minority status, automatic adjustment was made for the scale score at baseline.
h Tobit regression model (scale score defined as censored from below), estimated with WLSMV. In addition to the adjustments for patient gender, age, and racial/ethnic minority status, automatic adjustment was made for the scale score at baseline.

Table G.5. Interaction Between Randomization Group and Patients' Ratings of Their Clinician's Quality of Communication in Associations With Study Outcomes[a]

Outcome	No. of Patients/ No. of Clinicians[b]	Tx Group[c]		Quality of Care[d]		Interaction[e]	
		β	*p*	β	*p*	β	*p*
Aim 1: Events at Target Visit							
Occurrence of discussion[f]	492/124	2.020	**< 0.001**	0.026	0.318	−0.012	0.743
Aim 2: Concordance at 3 Months[g]	289/116	0.549	0.076	0.008	0.745	0.015	0.672
Aim 3: Depression and Anxiety							
Standard PHQ-8 score, 3 months[h]	488/124	0.342	0.398	0.025	0.436	0.026	0.543
Standard GAD-7 score, 3 months[i]	490/124	0.088	0.857	9.476	0.357	−16.100	0.261
Standard PHQ-8 score, 6 months[h]	487/124	0.313	0.500	0.024	0.469	−0.040	0.477
Standard GAD-7 score, 6 months[i]	490/124	−0.167	0.764	6.443	0.533	7.906	0.625

a Results were based on complex regression models with patients clustered under clinicians, estimated with restricted maximum likelihood. Each model included 3 predictors (randomization group, QOC latent construct, and the product of those 2 variables), with automatic adjustment for patient age, gender, and racial/ethnic minority status.

b Number of patients/number of clinician clusters. These sample sizes are larger than those shown for the main analyses for the study because they include patients whose only contribution was to the construction of the latent QOC variable.

c 0 = control; 1 = intervention.

d Quality of communication was a continuous latent variable measured at baseline with the 4 communication ratings judged to have been best supported by the intervention and without constraints designed to produce invariance between treatment groups or over time. Because of large floor effects, the 4 indicators were designed as censored from below.

e Term computed as the product of the randomization group indicator (0 = control, 1 = intervention) and the QOC latent construct.

f Binary outcome (0 = no discussion, 1 = discussion occurred).

g Binary outcome (1 = treatment preference and actual treatment at 3 months were both life extension or comfort care; 0 = treatment preference at 3 months was life extension and actual treatment was comfort care, or the reverse; or patient wasn't sure about preference or actual treatment). In addition to the automatic adjustments for patient gender, age, and racial/ethnic minority status, adjustment was made for treatment preference at 3 months (life extension or comfort care); patients with other values on this adjustment variable were excluded.

h Continuous variable, estimated with robust linear regression. In addition to the automatic adjustments for patient gender, age, and racial/ethnic minority status, adjustment was made for the standard Patient Health Questionnaire scale score at baseline.

i Variable defined as censored from below. In addition to the automatic adjustments for patient gender, age, and racial/ethnic minority status, adjustment was made for the standard generalized anxiety disorder scale score at baseline.