Appendix

1 Selection of eligible patients

Subjects for this study were eligible if they had one of six life-limiting conditions (heart failure, COPD, AIDS/HIV, selected neurodegenerative, renal failure, liver failure) and excluded if their primary diagnosis indicated trauma or if they received a transplant during their admission. Conditions, traumas and transplant procedures were identified through ICD-9 codes from the hospital databases. Specific ranges of ICD-9 codes for these factors are provided in Appendix Table 1.

Appendix Table 1 ICD-9 codes for identifying conditions, trauma and transplant in defining the sample

Conditions	
Heart failure	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.x
COPD	416.8, 416.9, 490.x – 505.x, 506.4, 508.1, 508.8
AIDS/HIV	042.x, 043.x, 044.x
Neurodegenerative	290.x, 294.1x, 294.2x, 330x – 337x
Renal	070.22, 070.23, 070.32, 070.33, 070.44, 070.54, 070.6, 070.9, 456.0-456.2, 570.x, 571.x, 572.2-572.8, 573.3, 573.4, 573.8,573.9, V42.7
Liver	403.01, 403.11, 403.91, 404.02, 404.12, 404.92, 585.x, 586.x, 588.0, V42.0, V45.1x, V56.x
Trauma	
As primary dx	348.1, 800.x-904.x, 925.x-929.x, 940.x-959.x, 994.0, 994.1
Transplant	
Bone marrow / stem cell:	41.00 – 41.09
Heart:	37.50 – 37.59
Lung:	33.50 – 33.59
Heart-lung:	33.60 – 33.69
Kidney:	55.60 – 55.69, 52.80
Liver:	50.50 – 50.59

2 Propensity score balance

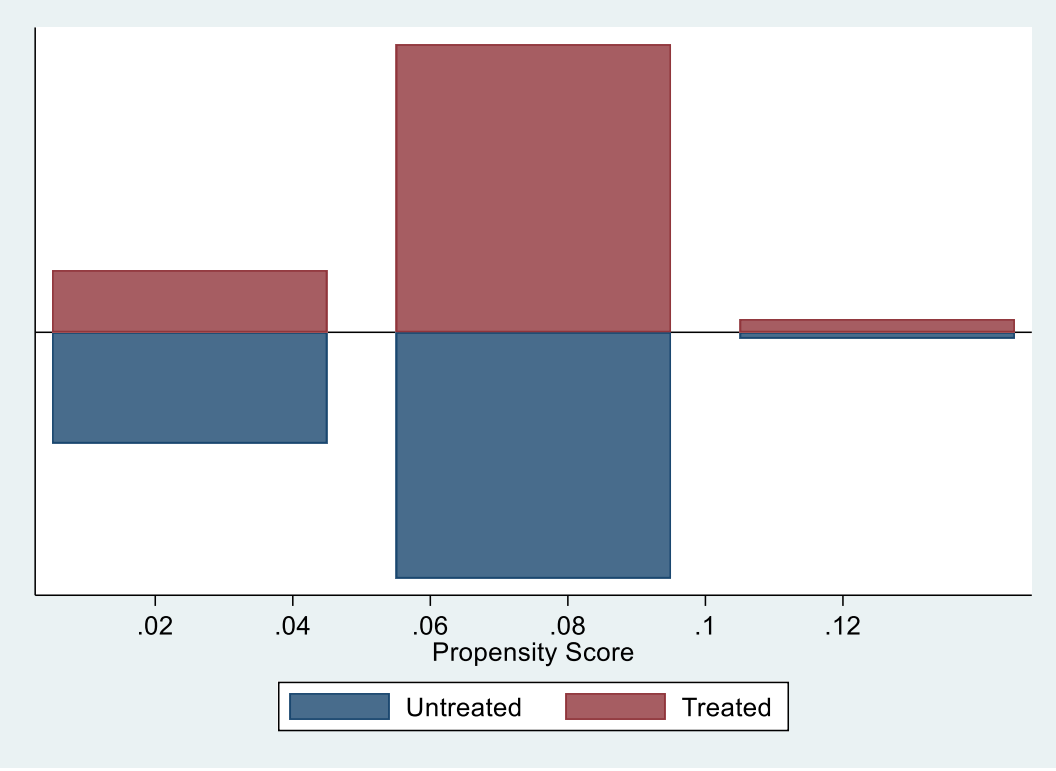
We constructed propensity scores sequentially, prior to estimating association between treatment and outcome, and with reference to "best practice" guidelines in this area.[1 2]

For each analytic sample defined by primary diagnosis (and in secondary analysis also multimorbidity count), we generated a sample-specific propensity score and followed the same process each time.[3]

We first identified observed covariates in the routinely collected data that were available in all datasets, collected at baseline (hospital admission) and hypothesized to be associated with treatment and outcome. These variables are listed in Table 2 of the main manuscript and were a fixed list for all analytic sub-samples. No recategorization or transformation (e.g. log of age) was necessary.

Second we assessed common support subjectively by checking overlap and similarity of distribution in the treatment and comparison groups. For example, here is the common support graph for those with a neurodegenerative diagnosis:

Appendix Figure 1 Common support for propensity score in group with neurodegenerative diagnosis (N=3317)



We repeated this for each propensity score created and in each case we considered the overlap and distribution to be satisfactory.

Third, we checked the balance of covariates within propensity score blocks. That is, we split the analytic sample into tertiles across the distribution of the propensity score and calculated absolute standardized differences for each covariate within each block.

Absolute standardized difference (ASD) is a measure of difference between groups that takes into account distribution (continuous variables) or prevalence (binary variables) and is not contingent on sample size. Formulae for calculating these are widely available elsewhere, e.g. [4].

No set guidelines exist for an acceptable level of difference but advice variously suggests 10-25% as satisfactory at this stage in the process.[1]

As an exemplar, ASD for each of the six variables across tertiles in the neurodegenerative group are presented below. Those ASDs between 10% and 25% are highlighted in orange; no ASD was over 25%. Across the 18 ASDs, median ASD is 6% and mean ASD is 8%. We take this to be acceptable. Our other samples did not differ substantively.

Appendix Table 2 Balance of covariates across the propensity score after weighting for those with neurodegenerative diagnosis (N=3317)

	Tertile 1					Tertile 2					Tertile 3				
	Comparison group N=1,065		Treatment group N=41			Comparison group N=1,039		Treatment group N=67			Comparison group N=1,020		Treatment group N=85		
	Mean	SD	Mean	SD	ASD	Mean	SD	Mean	SD	ASD	Mean	SD	Mean	SD	ASD
Age	58.06	16.49	58.51	14.98	3%	76.23	10.67	78.43	7.72	24%	82.39	9.29	81.75	10.52	6%
Male	0.47		0.46		2%	0.45		0.48		6%	0.38		0.35		5%
Medicaid	0.39		0.44		10%	0.95		0.97		12%	0.99		0.98		<1%
Medicare	0.11		0.10		4%	0.00		0.00		0%	0.00	0.00	0.00	0.00	0%
Elixhauser	2.13	1.55	2.32	1.51	13%	2.08	1.29	1.87	1.18	17%	1.94	1.39	2.01	1.34	5%
Walraven	2.02	4.60	2.24	4.61	5%	2.87	4.54	2.19	3.52	17%	6.67	5.92	7.32	6.20	11%

SD: standard deviation; ASD: absolute standardized difference. See Table 3 in main manuscript for details of predictors.

Fourth, after weighting the sample using inverse propensity score weights, we assessed balance in the sample.

We calculated absolute standardized difference for each covariate across treatment and comparison group before and after weighting. No set guidelines exist for an acceptable level of difference but advice variously suggests 10% as satisfactory at this stage in the process.[1]

The example of neurodegenerative diagnosis is provided below with values over 10% highlighted in red. Prior to weighing, four of the six covariates exhibited a high level of imbalance. After weighting, level of balance was very high (ASD<0.00001% in all cases). Other groups exhibited a similar level of balance after weighting, which is the typical result of the R program.[5]

Appendix Table 3 Balance of covariates before and after weighting in group with neurodegenerative diagnosis (N=3317)

	Unweighted					Weighted				
	Comparison group		Treatment group			Comparison group		Treatment group		
	Mean	**SD**	**Mean**	**SD**	**ASD**	**Mean**	**SD**	**Mean**	**SD**	**ASD**
Age	70.81	17.36	75.66	14.05	31%	75.66	14.48	75.66	14.05	<1%
Male	0.44	0.50	0.42	0.49	4%	0.42		0.42		<1%
Medicaid	0.73		0.86		34%	0.86		0.86		<1%
Medicare	0.10		0.02		35%	0.02		0.02		<1%
Elixhauser	2.03	1.41	2.03	1.33	<1%	2.03	1.39	2.03	1.33	<1%
Walraven	3.75	5.32	4.46	5.65	13%	4.46	5.62	4.46	5.65	<1%

SD: standard deviation; ASD: absolute standardized difference. See Table 3 in main manuscript for details of predictors.

3 Regression output

We present below the regression output for other predictors from our six primary cost analyses (Table 3 in the main manuscript). Please note that these coefficients and p values are open to misinterpretation. The purpose of the propensity score weights is to balance the treatment and comparison groups on observed confounders and so isolate best estimate of treatment 'effect' on outcome. The coefficient and significance of predictors on which the groups have been weighted is not a good indicator of each predictor's true association with outcome. Finally, average treatment effect estimates in the primary analysis are derived in dollars using bootstrapping and the -margins- command in Stata; the coefficients from routine regression output do not represent $ associations.

Appendix Table 4 Regression output from primary costs analysis for patients with HEART DISEASE diagnosis (N=28174)

	Coefficient	**Robust Standard error**	**Z**	**P value**	**95% CI lower**	**95% CI upper**
Received PC	-0.27	0.05	-6.07	<0.01	-0.36	-0.19
Age	-0.02	0.00	-9.17	<0.01	-0.03	-0.02
Male	0.02	0.04	0.45	0.65	-0.07	0.11
Medicaid	-0.18	0.12	-1.47	0.14	-0.42	0.06
Medicare	-0.01	0.09	-0.11	0.91	-0.19	0.17
Elixhauser	0.04	0.02	2.15	0.03	0.00	0.08
Walraven	0.03	0.00	6.45	<0.01	0.02	0.04
_cons	10.25	0.17	60.65	<0.01	9.92	10.58

See Table 3 in main manuscript for details of predictors. CI: confidence interval.

Appendix Table 5 Regression output from primary costs analysis for patients with NEURODEGENERATIVE diagnosis (N=3317)

	Coefficient	Robust Standard error	Z	P value	95% CI lower	95% CI upper
Received PC	-0.50	0.06	-8.35	<0.01	-0.62	-0.38
Age	-0.02	0.00	-6.35	<0.01	-0.03	-0.02
Male	-0.02	0.06	-0.38	0.71	-0.14	0.10
Medicaid	-0.16	0.24	-0.64	0.53	-0.63	0.32
Medicare	-0.34	0.12	-2.97	<0.01	-0.57	-0.12
Elixhauser	0.08	0.03	3.19	<0.01	0.03	0.13
Walraven	0.01	0.01	1.57	0.12	0.00	0.03
_cons	10.62	0.20	52.60	<0.01	10.23	11.02

See Table 3 in main manuscript for details of predictors. CI: confidence interval.

Appendix Table 6 Regression output from primary costs analysis for patients with COPD diagnosis (N=22747)

	Coefficient	Robust Standard error	Z	P value	95% CI lower	95% CI upper
Received PC	-0.24	0.05	-4.53	<0.01	-0.34	-0.14
Age	-0.01	0.00	-6.54	<0.01	-0.02	-0.01
Male	0.18	0.05	3.45	<0.01	0.08	0.28
Medicaid	0.09	0.13	0.70	0.48	-0.17	0.35
Medicare	0.08	0.08	1.07	0.28	-0.07	0.24
Elixhauser	0.07	0.02	3.53	<0.01	0.03	0.11
Walraven	0.01	0.00	2.82	0.01	0.00	0.02
_cons	-0.24	0.05	-4.53	<0.01	-0.34	-0.14

See Table 3 in main manuscript for details of predictors. CI: confidence interval.

Appendix Table 7 Regression output from primary costs analysis for patients with KIDNEY FAILURE diagnosis (N=6382)

	Coefficient	Robust Standard error	Z	P value	95% CI lower	95% CI upper
Received PC	-0.35	0.08	-4.40	<0.01	-0.51	-0.20
Age	-0.02	0.00	-7.02	<0.01	-0.03	-0.01
Male	0.14	0.09	1.52	0.13	-0.04	0.33
Medicaid	-0.34	0.32	-1.06	0.29	-0.97	0.29
Medicare	-0.33	0.27	-1.20	0.23	-0.86	0.21
Elixhauser	0.00	0.03	0.09	0.93	-0.05	0.05
Walraven	0.02	0.01	3.43	<0.01	0.01	0.04
_cons	10.42	0.38	27.52	<0.01	9.68	11.17

See Table 3 in main manuscript for details of predictors. CI: confidence interval.

Appendix Table 8 Regression output from primary costs analysis for patients with AIDS/HIV diagnosis (N=3068)

	Coefficient	Robust Standard error	Z	P value	95% CI lower	95% CI upper
Received PC	-0.12	0.09	-1.31	0.19	-0.30	0.06
Age	-0.01	0.00	-1.63	0.10	-0.02	0.00
Male	0.02	0.11	0.14	0.89	-0.20	0.23
Medicaid	-0.62	0.23	-2.67	0.01	-1.07	-0.16
Medicare	-0.67	0.23	-2.88	<0.01	-1.12	-0.21
Elixhauser	0.09	0.03	3.12	<0.01	0.03	0.14
Walraven	0.02	0.01	3.30	<0.01	0.01	0.03
_cons	9.38	0.32	29.35	<0.01	8.76	10.01

See Table 3 in main manuscript for details of predictors. CI: confidence interval.

Appendix Table 9 Regression output from primary costs analysis for patients with LIVER FAILURE diagnosis (N=9616)

	Coefficient	Robust Standard error	Z	P value	95% CI lower	95% CI upper
Received PC	-0.55	0.08	-7.28	<0.01	-0.70	-0.40
Age	-0.01	0.00	-3.38	<0.01	-0.02	0.00
Male	0.14	0.07	1.95	0.05	0.00	0.29
Medicaid	-0.48	0.12	-3.90	<0.01	-0.72	-0.24
Medicare	-0.58	0.10	-5.54	<0.01	-0.78	-0.37
Elixhauser	0.01	0.03	0.37	0.71	-0.05	0.07
Walraven	0.04	0.01	7.57	<0.01	0.03	0.05
_cons	9.76	0.21	46.50	<0.01	9.35	10.17

See Table 3 in main manuscript for details of predictors. CI: confidence interval.

4 Sensitivity analyses

Controlling for observed confounding using propensity scores may in some cases exacerbate unobserved biases and otherwise influence results.[6] We re-ran our primary analyses without propensity scores to check robustness to this aspect of analysis. The results are presented in Appendix Table 10. Our primary conclusions are unchanged:

- Results from evaluation of the association between treatment and outcome for each of the six diagnostic groups are consistent with the primary analysis in the manuscript. Treatment is associated with a statistically significant reduction in both total direct costs and hospital days for five of six conditions, the exception being AIDS/HIV;
- ANOVA statistic for these six estimates on costs, is again significant.
- ANOVA statistic for these six estimates on LOS is now significant (p=0.04) where in the primary analysis it was not (p=0.05).
- Post-hoc tests of head-to-head difference (Table 4 in the main manuscript; sensitivity analyses not shown) exhibit three differences to the primary analysis:
    - For costs, liver versus neurodegenerative, and liver versus kidney are not significant without propensity score weights.
    - For LOS, neurodegenerative versus COPD is significant without propensity score weights.

In summary, we consider our primary analyses to be substantively supported by this sensitivity analysis. All 12 analyses (six diagnostic groups, two outcomes each) are substantively consistent. Of 30 secondary head-to-head comparisons (15 per outcome of interest) of these estimates, 27 are substantively consistent.

Appendix Table 10 Rerun of primary analyses (Table 3 in main manuscript) without propensity scores: Estimated treatment effects on direct costs (USD) and LOS (days), by primary diagnosis

				Total direct costs ($)			LOS (days)		
Diagnosis	**All (N=)**	**CG (n=)**	**TG (n=)**	**ATET ($)**	**95% CI**	**One-way ANOVA**	**ATET (days)**	**95% CI**	**One-way ANOVA**
Heart failure	28174	27340	834	**-2666**	**-3440 to -1892**		**-1.20**	**-1.77 to -0.64**	
Neurodegenerative	3317	3124	193	**-3523**	**-4394 to -2651**		**-2.76**	**-3.40 to -2.12**	
COPD	22747	22332	415	**-1613**	**-2217 to -1009**	**F(5, 1953)= 9.5,**	**-1.13**	**-1.66 to -0.60**	F(5, 1953)= 2.2,
Kidney failure	6382	6226	156	**-3589**	**-5132 to -2045**	**p<0.0005**	**-2.32**	**-3.18 to -1.46**	p=0.05
HIV/AIDS	3068	2944	124	-2564	-6311 to 1184		-1.07	-2.97 to 0.84	
Liver failure	9616	9379	237	**-7574**	**-9232 to -5916**		**-1.57**	**-2.36 to -0.78**	

TG: Treatment group, receiving palliative care within three days of admission; CG: comparison group, including all other subjects. ATET: average treatment effect on the treated. CI: confidence interval.

References

1. Garrido MM, Kelley AS, Paris J, et al. Methods for constructing and assessing propensity scores. Health services research 2014;**49**(5):1701-20.
2. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in medicine 2015;**34**(28):3661-79.
3. Green KM, Stuart EA. Examining moderation analyses in propensity score methods: application to depression and substance use. Journal of consulting and clinical psychology 2014;**82**(5):773-83.
4. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in medicine 2009;**28**(25):3083-107.
5. Imai K, Ratkovic M. Covariate balancing propensity score. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2014;**76**(1):243-63.
6. Brooks JM, Ohsfeldt RL. Squeezing the balloon: propensity scores and unmeasured covariate balance. Health services research 2013;**48**(4):1487-507.