Supplementary Materials, Figures

Table S1: Description of treatment episodes follow-up (14-180 days after treatment initiation)

	All (n=5,554)	KPC (n=2,559)	KPW (n=2,995)
# visits, total	24,189	10,305	13.884
Median # (IQR[1]) visits	3 (2, 6)	3 (1, 5)	3 (2, 6)
Median # (IQR) PHQ-9 scores[2]	2 (1, 5)	2 (1, 4)	3 (1, 5)
# (%) of visits with PHQ-9	19, 048 (79%)	7,608 (74%)	11,440 (82%)
Median (IQR) days since baseline of final visit	104 (49, 154)	87 (41, 147)	118 (58, 160)
Median (IQR) days since baseline of final PHQ-9	93 (42, 148)	81 (36, 141)	104 (49, 154)
Mean (sd[3]) best PHQ-9 score	9.6 (5.9)	9.4 (6.0)	9.7 (5.9)
Mean (sd) finalPHQ-9 score	11.0 (6.4)	10.8 (6.40	11.3 (6.4)
# (%) best = final PHQ-9 score	3,703 (67%)	1,812 (71%)	1,891 (63%)

[1]IQR: Interquartile range

[2]Some encounters had more than 1 PHQ-9 observations

[3]sd: standard deviation

Table S2(a-l): Cross-tabulation of depression treatment outcome measures

Table S2(a): KPC Response vs. Remission

	No remission			Remission			Total
	Count	Row %	Col %	Count	Row %	Col %	
No response	1322	100%	66.70%	0	0%	0%	1322
Response	660	53.40%	33.30%	577	46.60%	100%	1237
Total	1982			577			2559

Table S2(b): KPC Response vs. Effect size ≥ 0.8

	Effect size < 0.8			Effect size ≥ 0.8			Total
	Count	Row %	Col %	Count	Row %	Col %	

No response	683	51.70%	100%	639	48.30%	34.10%	1322
Response	0	0%	0%	1237	100%	66%	1237
Total	683			1876			2559

Table S2(c): KPC Response vs. SAES ≥ 0.8

	SAES < 0.8			SAES ≥ 0.8			Total
	Count	Row %	Col %	Count	Row %	Col %	
No response	818	61.90%	100%	504	38.10%	28.90%	1322
Response	0	0%	0%	1237	100%	71.10%	1237
Total	818			1741			2559

Table S2(d): KPC Effect size ≥ 0.8 vs. Remission

	No remission			Remission			Total
	Count	Row %	Col %	Count	Row %	Col %	
Effect size < 0.8	683	100%	34.50%	0	0%	0%	683
Effect size ≥ 0.8	1299	69.20%	65.50%	577	30.80%	100%	1876
Total	1982			577			2559

Table S2(e): KPC SAES ≥ 0.8 vs. Remission

	No remission			Remission			Total
	Count	Row %	Col %	Count	Row %	Col %	
SAES < 0.8	818	100%	41.30%	0	0%	0%	818
SAES ≥ 0.8	1164	66.90%	58.70%	577	33.10%	100%	1741
Total	1982			577			2559

Table S2(f): KPC SAES ≥ 0.8 vs. Effect size ≥ 0.8

	Effect size < 0.8			Effect size ≥ 0.8			Total
	Count	Row %	Col %	Count	Row %	Col %	
SAES < 0.8	624	76.30%	91.40%	194	23.70%	10.30%	818
SAES ≥ 0.8	59	3.40%	8.60%	1682	96.60%	89.70%	1741

Total	683			1876			2559

Table S2(g): KPW Response vs. Remission

	No remission			Remission			Total
	Count	Row %	Col %	Count	Row %	Col %	
No response	1699	1	0.717	0	0	0	1699
Response	670	0.517	0.283	626	0.483	1	1296
Total	2369			626			2995

Table S2(h): KPW Response vs. Effect size $\geq$ 0.8

	Effect size < 0.8			Effect size $\geq$ 0.8			Total
	Count	Row %	Col %	Count	Row %	Col %	
No response	888	0.523	1	811	0.477	0.385	1699
Response	0	0	0	1296	1	0.615	1296
Total	888			2107			2995

Table S2(i): KPW Response vs. SAES $\geq$ 0.8

	SAES < 0.8			SAES $\geq$ 0.8			Total
	Count	Row %	Col %	Count	Row %	Col %	
No response	1049	61.70%	100%	650	38.30%	33.40%	1699
Response	0	0%	0%	1296	100%	67%	1296
Total	1049			1946			2995

Table S2(j): KPW Effect size $\geq$ 0.8 vs. Remission

	No remission			Remission			Total
	Count	Row %	Col %	Count	Row %	Col %	
Effect size < 0.8	888	100%	37.50%	0	0%	0%	888
Effect size $\geq$ 0.8	1481	70.30%	62.50%	626	29.70%	100%	2107
Total	2369			626			2995

Table S2(k): KPW SAES ≥ 0.8 vs. Remission

	No remission			Remission			Total
	Count	Row %	Col %	Count	Row %	Col %	
SAES < 0.8	1049	100%	44.30%	0	0%	0%	1049
SAES ≥ 0.8	1320	67.80%	55.70%	626	32.20%	100%	1946
Total	2369			626			2995

Table S2(l): KPW SAES ≥ 0.8 vs. Effect size ≥ 0.8

	Effect size < 0.8			Effect size ≥ 0.8			Total
	Count	Row %	Col %	Count	Row %	Col %	
SAES < 0.8	772	73.60%	86.90%	277	26.40%	13.10%	1049
SAES ≥ 0.8	116	6%	13.10%	1830	94%	87%	1946
Total	888			2107			2995

Table S3(a-b): Association between baseline PHQ-9 score and probability of treatment success, best PHQ-9 score

Table S3(a): Primary analysis, additive adjustment for site

	OR[1]	95% CI[2], lower	95% CI, higher	p-value
Effect size ≥ 0.8	2.306	2.01	2.649	<0.001
SAES ≥ 0.8	0.607	0.536	0.687	<0.001
Response	1.064	0.946	1.197	0.299
Remission	0.431	0.37	0.502	<0.001

[1] Odds ratio (OR) for 10-point increase in baseline score (e.g., 20 vs. 10 or 25 vs. 15). This OR is reported for all analyses.

[2] CI: confidence interval

Table S3(b): Interaction between site and probability of success

	KPC			KPW			
	OR	95% CI, lower	95% CI, higher	OR	95% CI, lower	95% CI, higher	p-value, interaction

Effect size ≥ 0.8	1.871	1.522	2.307	2.705	2.25	3.262	<0.001
SAES ≥ 0.8	0.591	0.488	0.716	0.618	0.525	0.728	<0.001
Response	1.092	0.914	1.304	1.044	0.892	1.221	0.544
Remission	0.408	0.324	0.511	0.452	0.367	0.554	<0.001

Table S4: Sensitivity analysis regression using inverse probability of observation weighting to account for loss to follow-up, *best* score

	OR	95% CI, lower	95% CI, higher	p-value
Effect size ≥ 0.8	2.269	2.024	2.547	<0.001
SAES ≥ 0.8	0.602	0.543	0.668	<0.001
Response	1.066	0.967	1.176	0.201
Remission	0.429	0.378	0.487	<0.001

Table S5(a-b): Sensitivity analysis using alternate methods to calculate effect size and severity-adjusted effect size (SAES), *best* score

Table S5(a): Effect size

	OR	95% CI, lower	95% CI, higher	p-value
Effect size ≥ 0.8, primary analysis	2.306	2.01	2.649	<0.001
Effect size ≥ 0.8, all baseline SD[1]	2.306	2.01	2.649	<0.001
Effect size ≥ 0.8, difference SD[2]	2.585	2.267	2.952	<0.001

[1] Effect size calculated using all baseline PHQ-9 scores from eligible episodes (including those without follow-up PHQ-9) to calculate standard deviation (SD).
[2] Effect size calculated using the standard deviation of the differences between best follow-up and baseline PHQ-9 score

Table S5(b): SAES

	OR	95% CI, lower	95% CI, higher	p-value

SAES ≥ 0.8, primary analysis	0.602	0.543	0.668	<0.001
SAES ≥ 0.8, IPW[1]	0.602	0.543	0.668	<0.001

[1] SAES calculates using inverse probability weighting (IPW) accounting for loss to follow-up for linear regression of expected best follow-up score given baseline

Table S6(a-c): Association between baseline PHQ-9 score and probability of treatment success varying the threshold that defines treatment success, *best* score

Table S6(a): Effect size

	OR	95% CI, lower	95% CI, higher	p-value
Effect size ≥ 0.6	2.387	2.056	2.776	<0.001
Effect size ≥ 0.8, primary analysis	2.306	2.01	2.649	<0.001
Effect size ≥ 1	2.603	2.282	2.973	<0.001
Effect size ≥ 1.2	2.912	2.563	3.312	<0.001

Table S6(b): SAES

	OR	95% CI, lower	95% CI, higher	p-value
SAES ≥ 0.6	0.537	0.471	0.613	<0.001
SAES ≥ 0.8, primary analysis	0.607	0.536	0.687	<0.001
SAES ≥ 1	0.709	0.629	0.799	<0.001
SAES ≥ 1.2	0.826	0.734	0.928	0.001

Table S6(c): Response

	OR	95% CI, lower	95% CI, higher	p-value
≥ 25% improvement	1.22	1.073	1.388	0.002
≥ 40% improvement	1.058	0.941	1.191	0.346
≥ 50% improvement, primary analysis	1.064	0.946	1.197	0.299
≥ 60% improvement	1.143	1.01	1.294	0.034

≥ 75% improvement	1.199	1.033	1.392	0.017

Table S7: Depression treatment outcomes, final PHQ-9 score

	All (n= 5,554)	**KPC** (n= 2,559)	**KPW** (n= 2,995)
Effect size ≥ 0.8	3,414 (61%)	1,638 (64%)	1,776 (59%)
SAES ≥ 0.8	2,545 (46%)	1,298 (68%)	1,247 (42%)
Response	2,026 (36%)	1,016 (40%)	1,010 (34%)
Remission	923 (17%)	440 (17%)	483 (16%)

Table S8(a-b): Association between baseline PHQ-9 score and probability of treatment success,
final PHQ-9 score

Table S8(a): Primary analysis, additive adjustment for site

	OR[1]	**95% CI**[2]**, lower**	**95% CI, higher**	**p-value**
Effect size ≥ 0.8	2.158	1.904	2.448	<0.001
SAES ≥ 0.8	0.785	0.698	0.884	<0.001
Response	1.085	0.961	1.225	0.189
Remission	0.451	0.381	0.534	<0.001

Table S8(b): Interaction between site and probability of success

	KPC			**KPW**			
	OR	**95% CI, lower**	**95% CI, higher**	**OR**	**95% CI, lower**	**95% CI, higher**	**p-value, interaction**
Effect size ≥ 0.8	1.949	1.611	2.361	2.329	1.973	2.754	<0.001
SAES ≥ 0.8	0.764	0.639	0.913	0.803	0.685	0.941	<0.001
Response	1.233	1.029	1.479	0.977	0.828	1.151	0.074
Remission	0.458	0.356	0.586	0.446	0.354	0.56	<0.001

Table S9: Sensitivity analysis regression using inverse probability of observation weighting to account for loss to follow-up, *final* score

	OR	95% CI, lower	95% CI, higher	p-value
Effect size ≥ 0.8	2.15	1.937	2.389	<0.001
SAES ≥ 0.8	0.785	0.711	0.867	<0.001
Response	1.097	0.991	1.214	0.073
Remission	0.452	0.393	0.52	<0.001

Table S10(a-b): Sensitivity analysis using alternate methods to calculate effect size and severity-adjusted effect size (SAES), *final* score

Table S10(a): Effect size

	OR	95% CI, lower	95% CI, higher	p-value
Effect size ≥ 0.8, primary analysis	2.158	1.904	2.448	<0.001
Effect size ≥ 0.8, all baseline SD[1]	2.158	1.904	2.448	<0.001
Effect size ≥ 0.8, difference SD[2]	2.555	2.261	2.889	<0.001

[1] Effect size calculated using all baseline PHQ-9 scores from eligible episodes (including those without follow-up PHQ-9) to calculate standard deviation (SD).
[2] Effect size calculated using the standard deviation of the differences between final follow-up and baseline PHQ-9 score

Table S10(b): SAES

	OR	95% CI, lower	95% CI, higher	p-value
SAES ≥ 0.8, primary analysis	0.785	0.711	0.867	<0.001
SAES ≥ 0.8, IPW[1]	0.788	0.714	0.869	<0.001

[1] SAES calculates using inverse probability weighting (IPW) accounting for loss to follow-up for linear regression of expected final follow-up score given baseline

Table S11(a-c): Association between baseline PHQ-9 score and probability of treatment success varying the threshold that defines treatment success, *final* score

Table S11(a): Effect size

	OR	95% CI, lower	95% CI, higher	p-value
Effect size ≥ 0.6	2.115	1.857	2.413	<0.001
Effect size ≥ 0.8, primary analysis	2.158	1.904	2.448	<0.001
Effect size ≥ 1	2.316	2.049	2.62	<0.001
Effect size ≥ 1.2	2.565	2.27	2.9	<0.001

Table S11(b): SAES

	OR	95% CI, lower	95% CI, higher	p-value
SAES ≥ 0.6	0.675	0.6	0.76	<0.001
SAES ≥ 0.8, primary analysis	0.785	0.698	0.884	<0.001
SAES ≥ 1	0.952	0.843	1.074	0.422
SAES ≥ 1.2	1.215	1.07	1.378	0.003

Table S11(c): Response

	OR	95% CI, lower	95% CI, higher	p-value
≥ 25% improvement	1.159	1.028	1.306	0.016
≥ 40% improvement	1.071	0.952	1.205	0.256
≥ 50% improvement, primary analysis	1.085	0.961	1.225	0.189
≥ 60% improvement	1.175	1.03	1.34	0.017
≥ 75% improvement	1.254	1.061	1.481	0.008

Figure S1: Flow diagram for identifying analytic cohort

Figure S1. Flow Diagram for Identification of the Cohort

Entire patient population with new psychotherapy starts

KPW N = 5,977 patients, n = 122 providers

KPCO N = 6,474 patients, n = 122 providers

Baseline PHQ-9 missing

KPW N = 260 patients

KPCO N = 1,042 patients

Baseline PHQ-9 available

KPW N = 5,717 patients, n = 115 providers

KPCO N = 5,432 patients, n = 118 providers

Baseline PHQ-9 < 10

KPW N = 1,752 patients

KPCO N = 1,312 patients

Baseline PHQ-9 ≥ 10

KPW N = 3,965 patients, n = 109 providers

KPCO N = 4,120 patients, n = 118 providers

No follow-up PHQ-9 between 2 weeks and 6 months after baseline visit

KPW N = 970 patients

KPCO N = 1,561 patients

One or more **follow-up PHQ-9** between 2 weeks and 6 months after baseline visit

KPW N = 2,995 patients, n = 105 providers

KPCO N = 2,559 patients, n = 113 providers

Figure S2: Distribution of baseline PHQ-9 scores for episodes with and without follow-up scores observed at Kaiser Permanente Colorado, KPC, (a) and Kaiser Permanente Washington, KPW, (b).

Figure S2(a) KPC

2,559 episodes with follow-up

1,561 episodes without follow-up

Figure S2(b) KPW

2,995 episodes with follow-up

970 episodes without follow-up
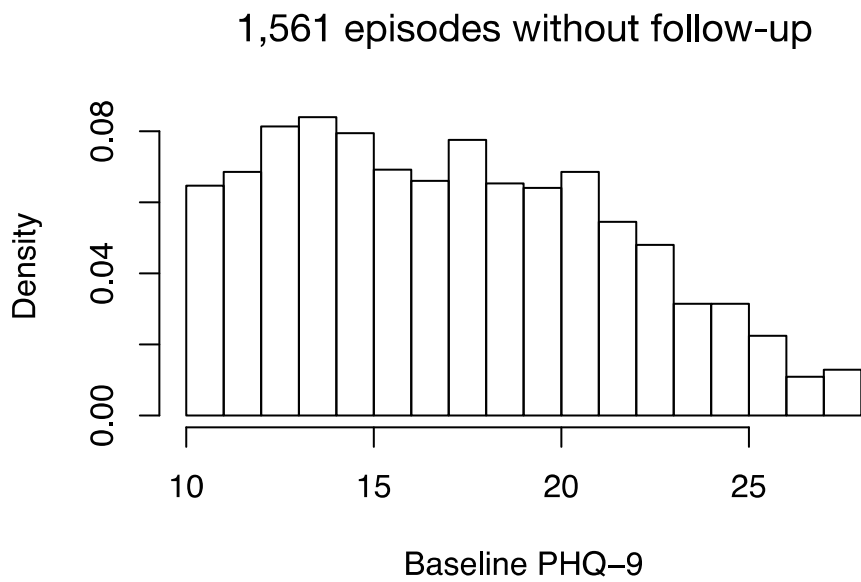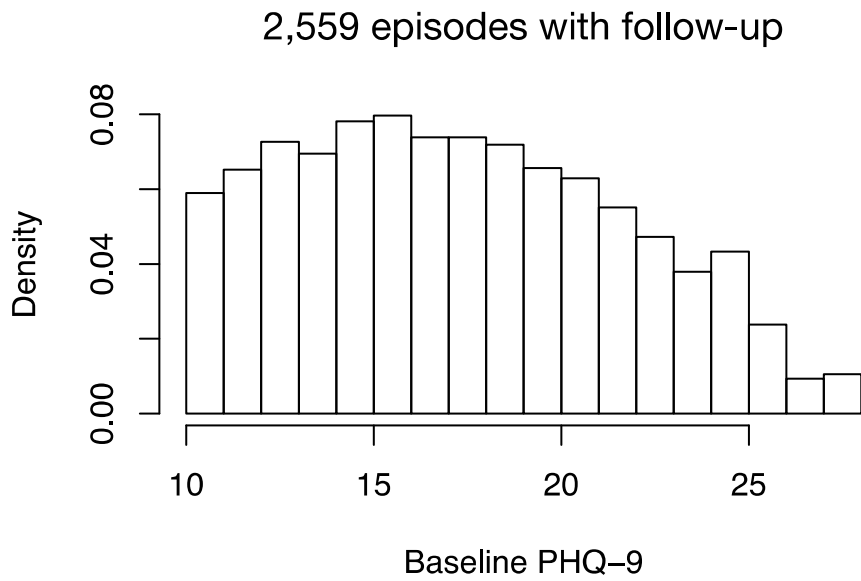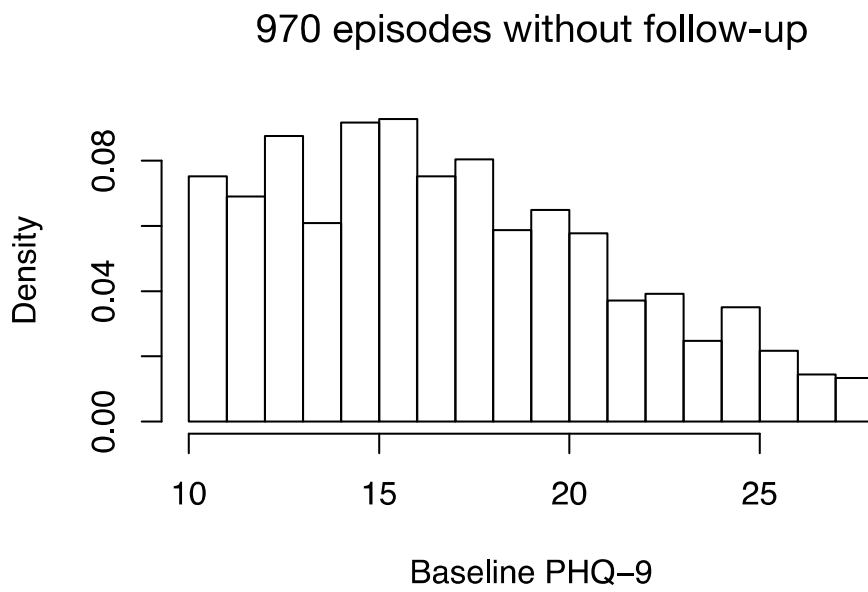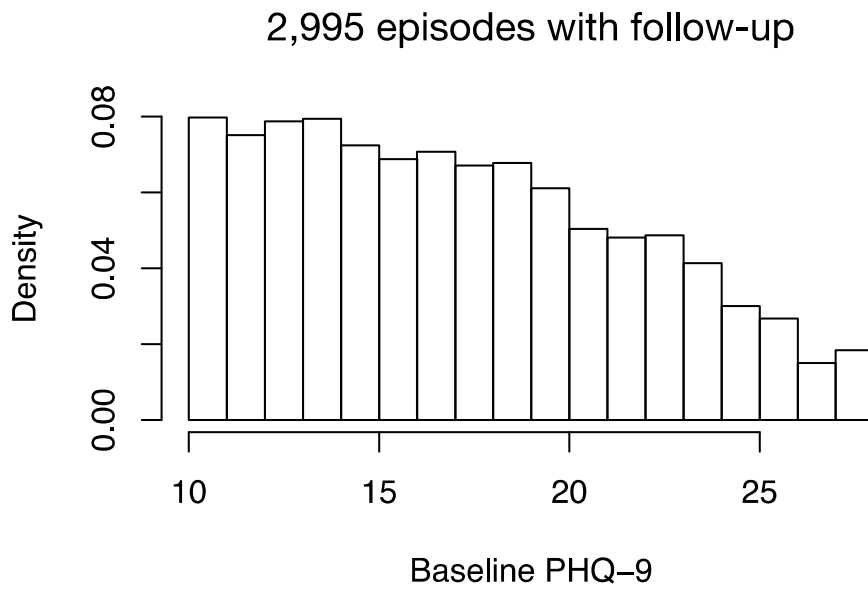
Figure S3: Distribution of best and final PHQ-9 scores 14-180 days after treatment initiation for new psychotherapy episodes at KPC (a) and KPW (b)

Figure S3(a) KPC

Best follow-up score

Follow-up PHQ−9

Final follow-up score

Follow-up PHQ−9

Figure S(b) KPW

Best follow-up score

Final follow-up score
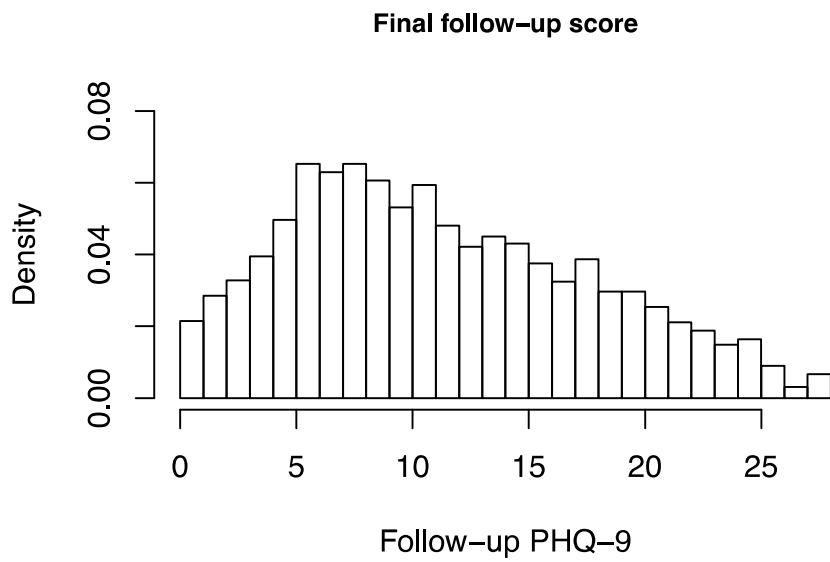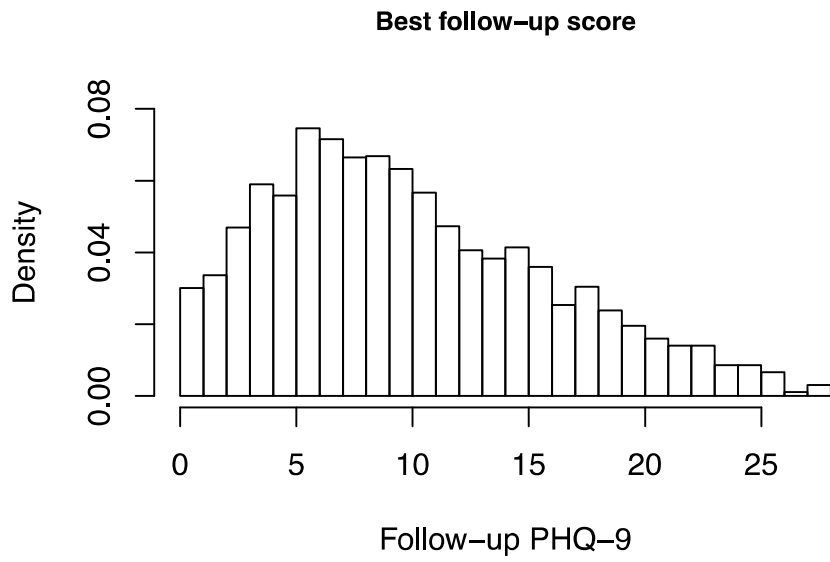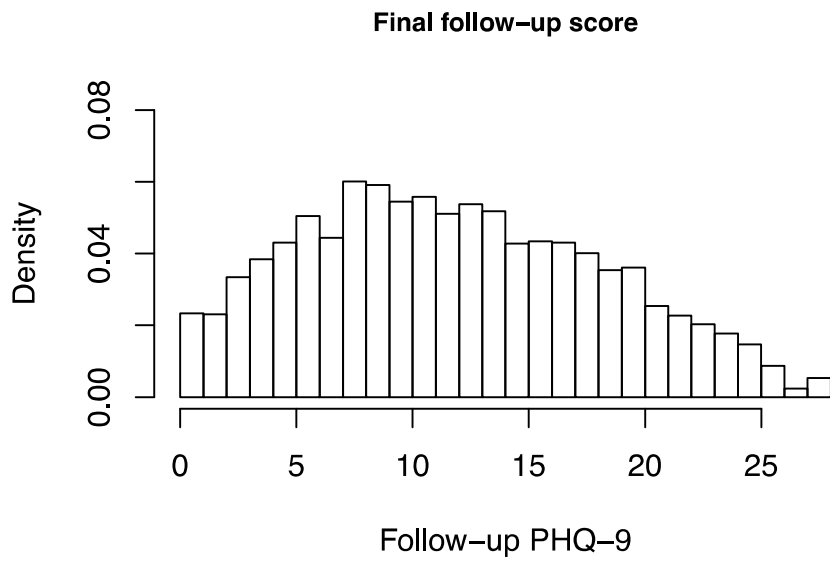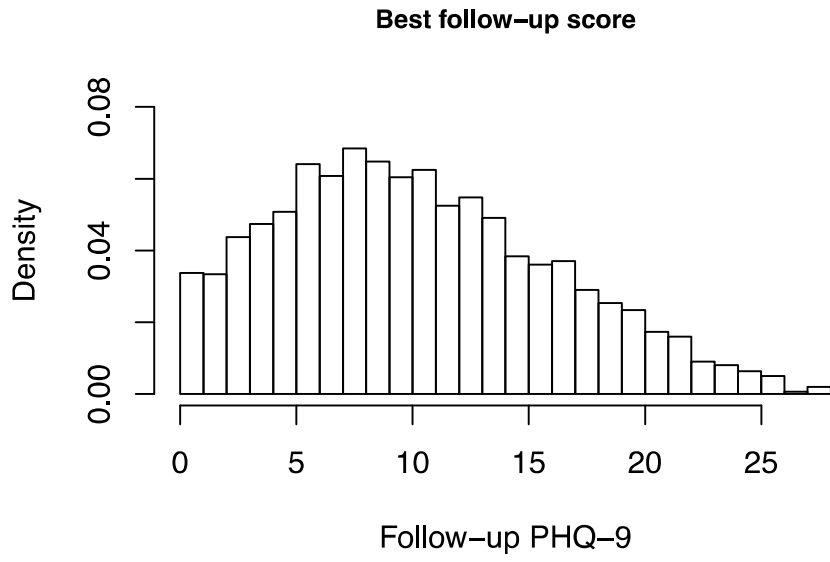
Figure S4: Probability of treatment success varying the threshold that defines treatment success, *effect size*, at KPC (a) and KPW (b)

Figure S4(a) KPC
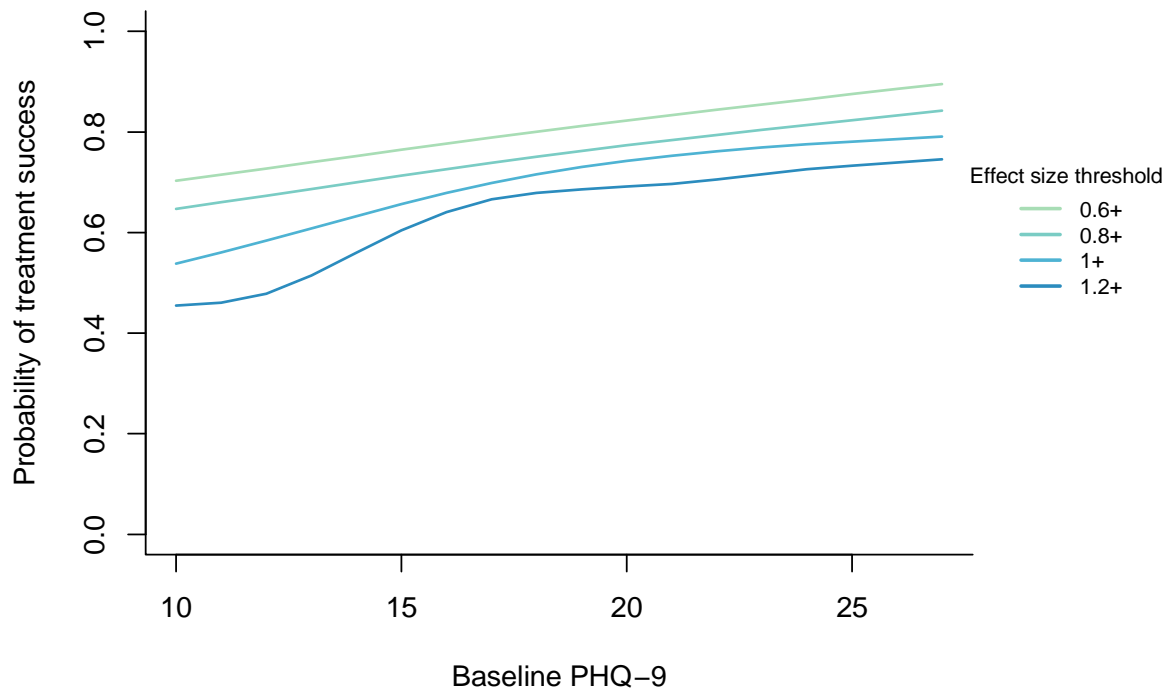
Figure S4(b)
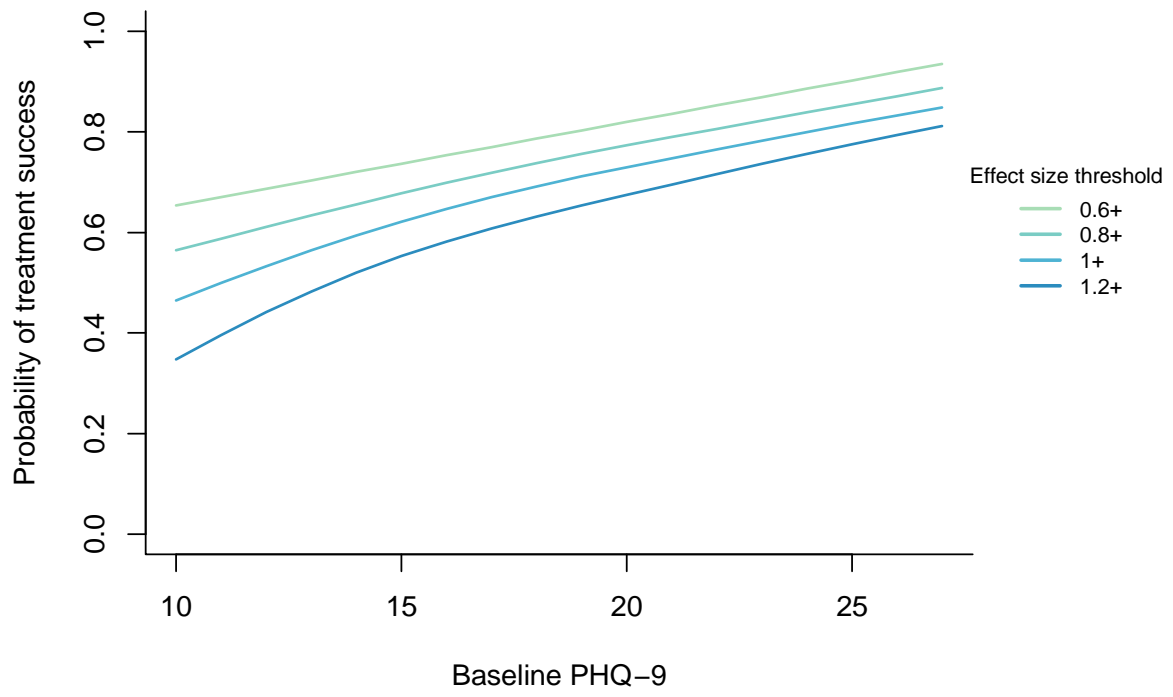
Figure S5: Probability of treatment success varying the threshold that defines treatment success, *SAES,* at KPC (a) and KPW (b)
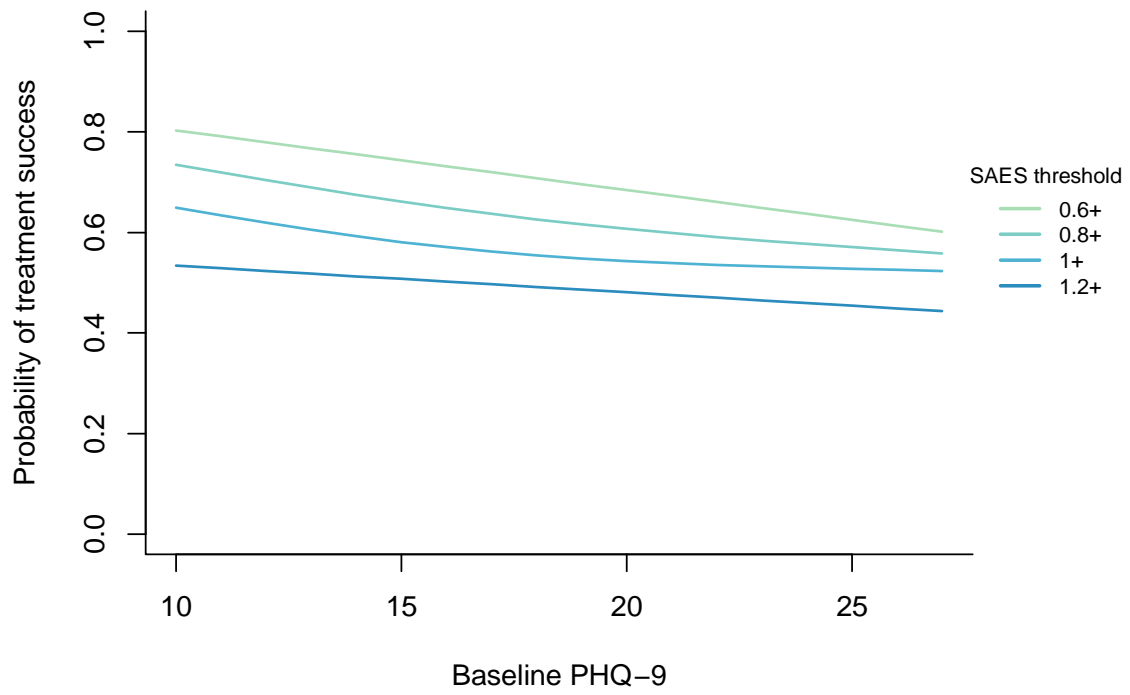
Figure S5(a) KPC
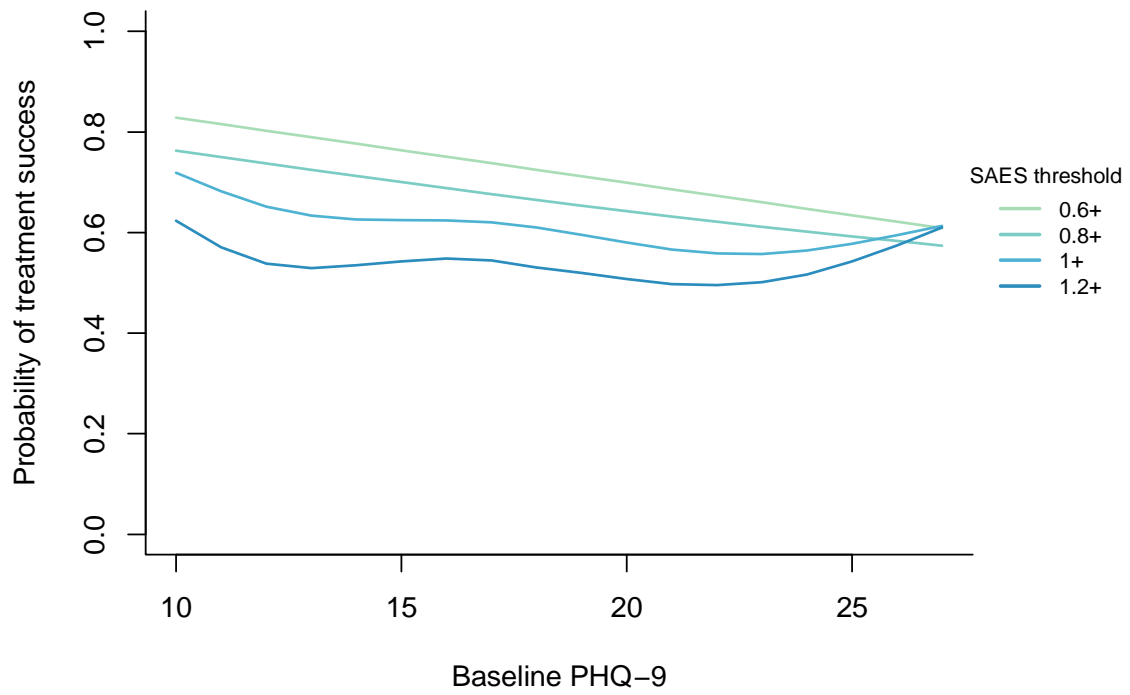
Figure S5(b) KPW

Figure S6: Probability of treatment success varying the threshold that defines treatment success, *response* (or percent improvement), at KPC (a) and KPW (b)
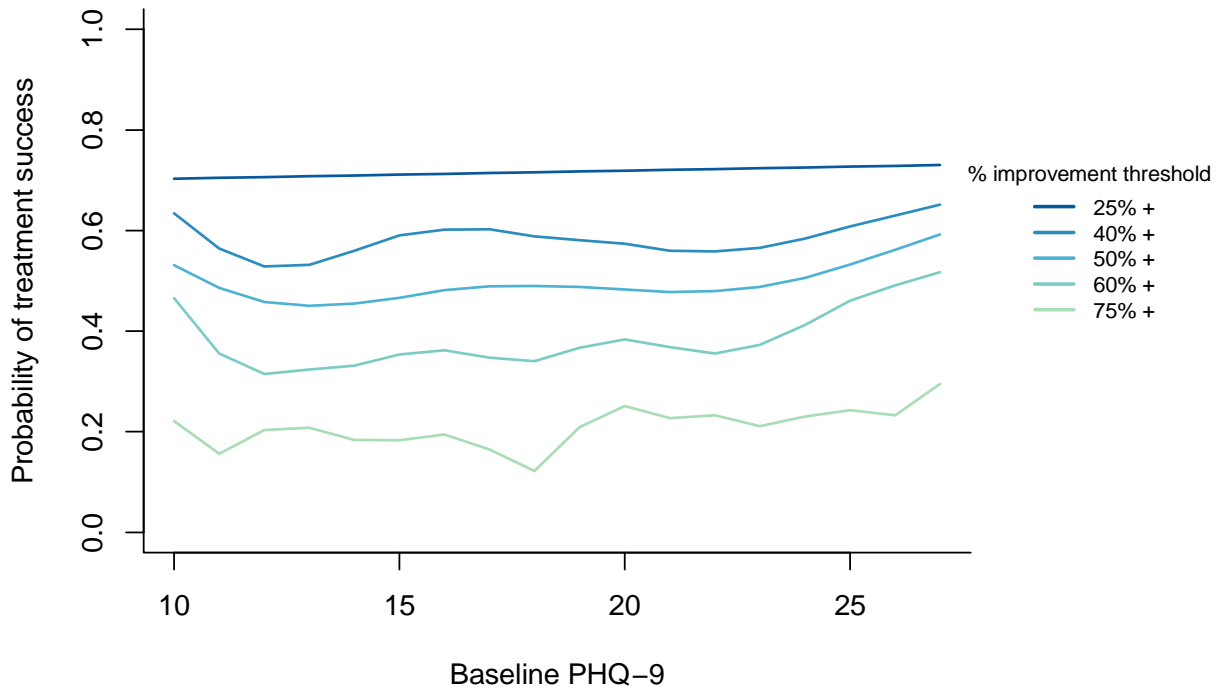
Figure S6(a) KPC

Figure S6(b)

KPW