

**SUMMARY.** We describe how to conduct a causal economic evaluation using mental health real-world data from Talking Therapies service data in England.



**OBJECTIVES.** Real-world evidence is playing an increasingly important role in health technology assessment, but is prone to selection and confounding bias. We demonstrate how to conduct a real-world within-study cost-per-quality-adjusted-life-year (QALY) analysis. We combined traditional within-trial bootstrapped regression-baseline-adjustment with causal inference methods, using a Target Trial (TT) framework, inverse probability weights (IPWs), marginal structural models (MSMs), and g-computation (g-comp), applied to England's Improving Access to Psychological Therapies (IAPT) mental-health e-records. IAPT is now formally referred to as Talking Therapies for anxiety and depression (TTad) services.

**METHODS.** The 'Assessing a Distinct IAPT service' (ADAPT) study evaluated an Enhanced-TTad service to account for the wider determinants of mental health (e.g., debt-based anxiety) against TTad's treatment-as-usual in a Geographical or Historical control. TTad collects patient-reported PHQ-9-depression and GAD-7-anxiety scores at index-assessment and each treatment session, from which we predicted EQ-5D utilities using a mapping function for QALY estimation.

We prespecified our TT including eligibility, treatment strategies, assignment procedure, follow-up, outcomes, estimands, and analysis plan (Table 1). We used stabilised treatment-related and censoring-related IPWs (sIPTWs and/or sIPCWs) within MSMs to reduce selection and confounding bias due to non-randomised treatment allocation and informative censoring, respectively. Our doubly-robust approach involved MSM-adjusted baseline covariates and g-computation to estimate incremental utilities, costs, and QALYs, with bootstrapped bias-corrected 95% confidence-intervals (95%bCIs) and CEACs.

**RESULTS.** Analysis sample: Enhanced, N=5,441; Geographical control, N=2,149. Naïve regression-baseline-adjustment and doubly-robust approaches (Tables 2&3) suggested Enhanced-TTad dominated treatment-as-usual, with average per-person (95%bCIs) cost-savings of £30.64 (£22.26 to £38.90) or £29.64 (£20.69 to £37.99) and QALYs-gained of 0.00035 (-0.00075 to 0.00152) or 0.00052 (-0.00105 to 0.00277), respectively; probability of cost-effectiveness at £30,000 per QALY was 99% or 95%, respectively. Doubly-robust and naïve results concurred; albeit, doubly-robust results suggested average QALY gains were higher but less certain. The cost-effectiveness results were driven by the potential for the Enhanced service to provide cost-savings. Cost-savings were supported by Historical comparisons, but incremental QALYs were uncertain.

**CONCLUSION.** When treatment allocation is non-randomised, the TT framework alongside doubly-robust analyses aids reduce selection and confounding bias.

Table 1. Target Trial protocol overview for the ADAPT study economic evaluation – analysis plan and secondary analyses not presented

Component	Description
Eligibility criteria	<b>New referrals to TTad-services:</b> no attendance at the TTad site in the previous 6-months since the new referral <b>Newly referred during:</b> March 2021 to March 2022 (intervention & geographical-control) or March 2018 and March 2019 (historical-control) <b>Baseline data:</b> recorded PHQ-9 (depression severity) and GAD-7 (anxiety severity) score at baseline – necessary for 'Condition caseness at baseline' <b>Baseline condition caseness:</b> classified as having depression caseness (PHQ-9≥10) or anxiety caseness (i.e., GAD-7≥8) at index assessment (i.e., baseline) <b>As-started treatment:</b> service-users had attended at least one treatment session to be defined as 'as-started' treatment <b>Sufficient TTad data for follow-up period:</b> available data time-horizon must be at least that of the analysis follow-up period (e.g., primary analysis: 16-weeks)
Treatment strategies	<b>Intervention:</b> Enhanced TTad-service (South-West, England), as TAU and/or 'Healthy Living Healthy Minds' programme and/or 1-1 Wellbeing Navigator sessions. <b>Geographical control:</b> TAU TTad-service in South-East, England; <b>Historical control:</b> TAU TTad-service in the intervention area but before the enhanced service had been implemented.
Assignment	<b>Non-randomised and unblinded:</b> all intervention site referrals are offered the enhanced TTad-service, with uptake based on service-user preference
Follow-up period(s)	<b>Baseline (time zero):</b> index appointment to assess condition caseness and allocate people to the waiting-list before first treatment session <b>Primary follow-up:</b> starts at baseline and ends at 16-weeks after baseline, regardless of TAU received and service discharge
Outcome(s)	<b>Primary:</b> PHQ-9 and GAD-7 scores, sex, and age mapped to EQ-5D-5L UK crosswalk utility scores for QALY estimation <b>Resource-use/costs:</b> TTad-service EHR recorded resources-use with costs applied for, or inflated to, the year 2020/21
Estimand(s) (causal contrasts)	<b>Primary ITT:</b> In new referrals to TTad-services, what is the between-group difference in mean TTad-service costs and QALYs accumulated since index assessment (i.e. baseline), with QALYs based on EQ-5D-5L crosswalk utilities predicted from PHQ-9 and GAD-7 scores, age, and sex, for those referred to the enhanced TTad-service compared to treatment-as-usual (TAU) for those within a geographical-control-site up to 16-weeks after baseline, regardless of TAU received and service discharge? <b>Primary PP:</b> same as 'as-started', but intervention group participants must have had at least one enhanced TTad-service treatment session
Acronyms & abbreviations	ADAPT, Assessing a Distinct Improving Access to Psychological Therapies (researchregistry7322); EHR, electronic health record; GAD-7, Generalized Anxiety Disorder-7 scale; HRQoL, health-related quality-of-life; ITT, intention-to-treat; PHQ-9, Patient Health Questionnaire-9 depression-scale; PP, per protocol; QALY, quality-adjusted life-year; TAU, treatment-as-usual.

Table 2. Cost-effectiveness results using different methods for ITT intervention vs geographical control over 16-weeks

No.	Method	Outcome	Intervention, N =5,441			Geographical, N=2,149			Average Treatment Effect (ATE)			
			Mean	95% CIs		Mean	95% CIs		Mean	95% CIs	ICER	
1	Naïve regression	QALYs	0.20532	0.20475	0.20589	0.20498	0.20402	0.20593	0.00035	-0.00078	0.00148	
	Naïve regression	Costs (£)	£173.85	£170.35	£177.36	£204.49	£197.29	£211.70	£-30.64	£-38.68	£-22.60	Dominates
2	MSM w/ sIPTW	QALYs	0.20555	0.20497	0.20613	0.20488	0.20395	0.20580	0.00067	-0.00043	0.00177	
	MSM w/ sIPTW	Costs (£)	£173.78	£170.28	£177.28	£203.42	£196.01	£210.84	£-29.64	£-37.82	£-21.47	Dominates
3	MSM w/ sIPCW	QALYs	0.20706	0.20651	0.20761	0.20685	0.20593	0.20778	0.00020	-0.00089	0.00130	
	No. 2	Costs (£)	£173.78	£170.28	£177.28	£203.42	£196.01	£210.84	£-29.64	£-37.82	£-21.47	Dominates
4	MSM w/ sIPTW*sIPCW	QALYs	0.20723	0.20668	0.20779	0.20676	0.20582	0.20769	0.00048	-0.00062	0.00157	
	No. 2	Costs (£)	£173.78	£170.28	£177.28	£203.42	£196.01	£210.84	£-29.64	£-37.82	£-21.47	Dominates
5	No. 4 & g-comp	QALYs	0.20519	N/A	N/A	0.20467	N/A	N/A	0.00052	N/A	N/A	
	No. 2 & g-comp	Costs (£)	£175.13	N/A	N/A	£204.77	N/A	N/A	£-29.64	N/A	N/A	Dominates

Table 3. Cost-effectiveness results for ITT and PP-intervention Vs control over 16-weeks with bootstrapped bias-corrected confidence intervals

Comparison	Method	QALYs			Costs, £			ICER	Prob. CE < λ per QALY		
		Mean	BC 95% bCIs		Mean	BC 95% bCIs			λ = £0	λ = £20k	λ = £30k
ITT (N=5,441) Vs. Geog (N=2,149)	Naïve regression	0.00035	-0.00075	0.00152	£-30.64	£-38.90	£-22.26	Dominates	100%	100%	99.3%
	Doubly robust	0.00052	-0.00105	0.00277	£-29.64	£-37.99	£-20.69	Dominates	100%	98.1%	95.0%
PP (N=549) Vs. Geog (N=2,149)	Naïve regression	-0.00478	-0.00729	-0.00233	£-34.09	£-45.95	£-22.70	< Q & < £	100%	49.2%	31.3%
	Doubly robust	-0.00157	-0.00736	0.00322	£-33.33	£-45.24	£-21.82	< Q & < £	100%	57.8%	46.2%
ITT (N=5,441) Vs. Hist (N=4,001)	Naïve regression	0.00364	0.00261	0.00471	£-84.54	£-92.78	£-76.61	Dominates	100%	100%	100%
	Doubly robust	0.00153	0.00025	0.00298	£-84.86	£-92.52	£-76.50	Dominates	100%	100%	100%
PP (N=549) Vs. Hist (N=4,001)	Naïve regression	0.00008	-0.00278	0.00262	£-85.50	£-97.60	£-75.07	Dominates	100%	99.9%	95.9%
	Doubly robust	-0.00103	-0.00535	0.00167	£-88.15	£-99.83	£-77.06	< Q & < £	100%	96.9%	89.4%