

Estimating number needed to treat (NNT) from continuous outcomes in randomised controlled trials: methodological challenges and a worked example using data from the UK Back Pain Exercise and Manipulation (BEAM) trial

Robert Froud,¹ Sandra Eldridge,¹ Ranjit Lall,² Martin Underwood²
¹Barts and the London School of Medicine and Dentistry
²Warwick Medical School

Background

- Using numbers needed to treat (NNT) to report trial outcomes can improve interpretability of RCT outcomes
- Authors of RCTs rarely report continuous outcomes using NNT, and rarely consider deteriorating patients
- Authors who do report NNTs, rarely report confidence intervals (CIs) and if they do, tend to use Wald intervals
- Wald intervals have poor coverage properties and may result in aberrations when used with NNT
- Recent consensus thresholds of 'responders to treatment' facilitate analyses of individual improvements and reporting of NNT in back pain trials
- The UK Back Pain Exercise and Manipulation (BEAM) trial was a large trial of physical therapies for LBP
- 1,334 patients consulting for LBP were randomised to either 'best' GP care, or 'best' GP care and exercise, manipulation or combined exercise and manipulation
- The primary endpoint was the change from baseline of the Roland Morris Disability Questionnaire (RMDQ) score at three and 12 months
- The trial showed a small to moderate mean difference between physical therapies and best 'GP' Care. Its clinical meaningfulness has been questioned. It's unclear how many patients would likely benefit from these treatments if they were prescribed
- Reporting the primary outcome of UK BEAM in terms of NNT may aid interpretation

Objectives

- To reanalyse the RMDQ data from UK BEAM, reporting results as NNT and to explore the effects of this reporting on interpretation of outcomes

Materials and methods

- Using consensus thresholds of 1) a five RMDQ point decrease to define improvement, and 2) a 30% change from baseline, we calculated NNTs for improvement
- Adopting equal and opposite magnitudes for thresholds of deterioration, allowed us to estimate 'Benefit' (additional improvements gained + additional deterioration prevented) at three and twelve months
- We used Newcombe's method 10 (Wilson scores) to estimate standard error, from which we derived CIs for the absolute risk difference and inverted these to obtain CIs for NNT point estimates
- For benefit we modified Newcombe's method 10 to incorporate extra variance introduced by our considering deterioration.

Results

Number needed to treat to gain, on average, one extra improvement or benefit compared to 'Best Care'

	Exercise		Manipulation		Combined	
<i>Three months</i>						
Improvement, 5 points	8.2	(4.9 to 25.0)	5.2	(3.7 to 8.8)	4.8	(3.5 to 7.8)
Improvement, 30%	9.0	(5.0 to 44.5)	5.4	(3.8 to 9.9)	4.4	(3.3 to 7.0)
Benefit, 5 points	6.6	(4.1 to 16.6)	5.1	(3.6 to 9.3)	4.3	(3.1 to 7.0)
Benefit, 30%	5.1	(3.4 to 10.7)	5.0	(3.4 to 9.8)	3.3	(2.5 to 4.9)
<i>12 months</i>						
Improvement, 5 points	*	-	8.4	(5.0 to 28.6)	9.0	(5.2 to 37.8)
Improvement, 30%	*	-	8.0	(4.9 to 24.3)	7.1	(4.5 to 17.9)
Benefit, 5 points	*	-	7.8	(4.5 to 29.0)	7.2	(4.3 to 22.8)
Benefit, 30%	*	-	6.9	(4.1 to 21.1)	5.8	(3.7 to 13.2)

* Analyses were not performed as no mean difference was reported between the exercise and best care groups at 12 months

Originally reported RMDQ outcomes of the UK BEAM trial

Group	Net benefit from intervention			
	at three months	(95% CI)	at 12 months	(95% CI)
Exercise	1.36**	(0.63–2.10)	0.39	(-0.41–1.19)
Manipulation	1.57***	(0.82–2.32)	1.01*	(0.22–1.81)
Combined treatment	1.87***	(1.15–2.60)	1.30**	(0.54–2.07)

Adapted from UK BEAM BMJ 2004;329:1377-81

* Significant at 5% level

** Significant at 1% level

*** Significant at 0.1% level

Discussion

- NNTs to achieve an improvement / benefit were small, ranging from two to nine
- These could be attractive to patients, clinicians, and purchasers
- Estimating NNT can be challenging due to difficulties in defining thresholds for individual responders to treatment on continuous scales, but recent consensus thresholds facilitates these analyses
- Where possible, NNT should be considered for future use alongside more conventional reporting methods

Thanks are due to Barts and the London charity for funding a PhD programme from which this work originates

r.j.froud@qmul.ac.uk

NNT software for STATA at :

www.robertfroud.info