Retrospective equivalence analysis of superiority trials: what are the implications for trial-based economic evaluations?

Author(s): David GT Whitehurst[1,2], James P Raftery[3] and Martyn Lewis[1]

1. Primary Care Musculoskeletal Research Centre, Keele University, Staffordshire.
2. Health Economics Facility, Birmingham University.
3. Wessex Institute for Health Research and Development, University of Southampton.

Contact details: d.whitehurst@cphc.keele.ac.uk

ABSTRACT

Aims: Many randomised clinical trials with active control groups (as opposed to placebo controls) fail to show superiority. This study investigates the implications for trial-based economic evaluations of performing retrospective equivalence analysis of superiority trials that have shown 'no difference' between interventions.

Methods: An equivalence analysis was performed on a pragmatic randomised clinical trial that compared the effectiveness of a brief pain management programme (BPM) and physical therapies (PT) for non-specific low back pain. This trial was originally designed as a superiority trial, and showed no significant differences between the interventions. A cost-utility analysis has been completed previously.

Data: Analysis was conducted using intention-to-treat and per-protocol approaches. We performed the analysis based on complete data and also using multiple imputation techniques to deal with missing data (four different analyses in total). Of the 402 patients recruited to the trial, 267 (66%) were defined as receiving their allocated treatment per-protocol. At 12 month follow-up, 329 (82%) patients provided data on the primary outcome measure.

Results: Using a retrospectively defined equivalence margin, results indicted the equivalence of the interventions in three of the four analyses, irrespective of the reduced statistical power. The CUA concluded that PT was a cost-effective primary care management strategy for low back pain. Conversely, adopting a CMA approach would suggest that BPM provides the more cost-effective use of health care resources.

Conclusions: Demonstrating the equivalence of alternative health care interventions is an insufficient justification for CMA, and could result in misleading conclusions concerning the cost-effectiveness of competing interventions.

Background

Pragmatic randomised clinical trials allow researchers to assess the effectiveness of health care interventions in routine clinical practice. Such trials attempt to reflect the heterogeneous patient groups that exist in reality (i.e. the eventual target population for an intervention). In theory, if an intervention is shown to provide a significant clinical benefit over an active control group (or vice versa), we also know that this clinical benefit would be maintained if the intervention were implemented in clinical practice.[1]

The aim of most randomised clinical trials (RCTs), whether pragmatic or explanatory, is to determine whether one intervention is superior to another (so-called 'superiority' trials). The key problem with superiority trials is that there are only two possible conclusions; either an intervention is superior, or not. An intervention that has been found to be 'not superior' falls short of showing that the alternative therapies are equivalent. An alternative is to perform an equivalence trial, where the aim is to demonstrate the therapeutic equivalence of two (or more) treatments. The choice of clinical trial design matters from a policy perspective, as demonstrating the equivalence of competing health care interventions would enable a choice to be offered to health care practitioners and/or patients. Given the large number of pragmatic RCTs where no significant differences are shown to exist, attention needs to be paid as to what conclusions can be drawn from these trials. Equivalence analysis on data from superiority trials may provide a valuable addition to clinical trial analysis, whether considered prospectively (i.e. incorporated in the trial design), or retrospectively.

The choice of clinical trial design may ultimately play a role in determining the appropriate form of analysis for a trial-based economic evaluation. It is widely accepted that an economic evaluation examining both the costs and consequences of the competing interventions should be performed alongside a superiority trial. For an equivalence trial, however, if the clinical effects are shown to be the same, only the costs of the interventions may be explored in a cost-minimisation analysis (CMA).[2]

This paper considers the implications for trial-based economic evaluations of performing equivalence analysis retrospectively using data from a superiority trial - a pragmatic RCT comparing alternative primary care management strategies for low back pain.[3] We start by describing the Keele low back pain trial, and further discuss the rationale for performing equivalence analysis on data from superiority trials. The following section reports a retrospective equivalence analysis performed on the

Keele trial data. Next we provide a brief summary of a cost-utility analysis performed alongside this trial.[4] Finally, we consider the implications for trial-based economic evaluations of performing equivalence analysis on data from superiority trials.

The Keele Low Back Pain Trial

Back pain in the UK poses a sizeable epidemiological and economic problem, with direct health care costs estimated to be £1.6 billion in 1998.[5] Guidelines for the management of acute low back pain in primary care recommend early referral to physiotherapy where first-line management by general practitioners (GPs) has not been effective, and emphasise the importance of recognising and addressing psychosocial factors known to be associated with long-term disability.[6,7] Evidence for the effectiveness of specific types of physiotherapy is conflicting, but supports active intervention above no treatment or sham procedures.[8,9] This trial assessed the effectiveness of a brief pain management programme (BPM), designed to target psychosocial factors, compared with physical therapy (PT) for patients consulting primary care with low back pain of less than 12 weeks' duration. Comprehensive details of the trial design, details of the interventions, and clinical findings have been published previously.[3]

Patients were randomly assigned to either BPM (n=201) or PT (n=201). Analysis was by intention to treat. The primary outcome measure was change in back-related disability. Disability was measured on the Roland and Morris Disability Questionnaire (RMDQ), a 24-item back-pain specific disability scale with scores ranging from 0 (best health) to 24 (worst health), expressed as changes between baseline and 12 month scores. There were no significant differences in RMDQ change scores between groups at 3 months or 12 months. For secondary outcomes, both groups were similar with respect to back pain and function, psychosocial measures and clinical assessments. A statistically significant difference was found regarding the number of treatment sessions, with the BPM package delivered in significantly fewer sessions (whether looking at the mean or median number of sessions).

The authors concluded that the BPM intervention "might be an efficient first-line approach to care of patients with low back pain presenting in primary-care practice".[3] This concluding statement is largely based on the two findings described above; there was insufficient evidence to show that one intervention was superior to the other, and the BPM package was delivered in significantly fewer sessions.

However, as this was a superiority trial, the concluding statement is preceded by the word 'might', as opposed to anything more definitive. From a policy perspective, the finding of no significant difference is disappointing. If the study had been able to conclude that BPM was equivalent to PT, as could be the conclusion of an equivalence trial, then policy makers would be able to offer both models of service. This could be important given the prevalence of back pain within the general public and the capacity constraints within primary care. Furthermore, additional factors such as patient and doctor preferences could be considered when choosing a back pain management strategy.[10] We performed a retrospective equivalence analysis on the same trial data to investigate the potential benefits of such an approach.

Retrospective Equivalence Analysis

Methods

There are a number of differences between superiority trials and equivalence trials concerning the design, conduct, and analysis of the study. For a thorough discussion of these issues, see Jones *et al*.[11] Two fundamental requirements for an equivalence trial are the setting of an equivalence margin, and the collection of data to identify whether individuals received their intervention according to the protocol of the trial.

Due to the inherent uncertainty in clinical studies, authors can never claim to have proven there is no difference between competing interventions. To determine whether interventions are therapeutically equivalent, it is necessary to define an *equivalence margin* – a range of values, defined *a priori*, within which the effect difference between interventions is considered to be clinically unimportant. If the confidence interval for the observed difference lies completely within this range, equivalence is demonstrated. Although setting the values for an equivalence margin is difficult, it allows researchers to be explicit in reporting the trial results.[12] The power calculation for the low back pain trial was based on a clinically significant difference of 2 points between treatment groups on the RMDQ. Although it is common for a clinically significant difference in an equivalence trial to be smaller than in a superiority trial, a change of 2 points on the RMDQ also forms the equivalence margin for this analysis (i.e. from –2 to +2). Evidence within the low back pain literature suggests that this would satisfy the requirement of 'a minimum difference of clinical importance'.[11,13]

In an equivalence trial it is recommended that both intention-to-treat (ITT) and per-protocol (PP) analyses should be performed, with the ideal scenario being the

same result in each case.[11] In the Keele trial, details of the techniques used within each treatment session were recorded on standard proformas, allowing each patient to be assessed for eligibility for a PP analysis. Performing the analysis retrospectively means that power calculations to determine sample size are uninformative. It is more appropriate to ascertain the power of our analyses given the sample size we have, e.g. given *n* number of patients in our ITT analysis, what power do we have to deem the interventions equivalent if they are, in fact, identical. These power calculations are reported in the results section.

As well as performing the analyses on observed data, i.e. patients that have responded to the primary outcome measure at baseline and 12 months, we have also used multiple imputation techniques to handle missing data. We generated 5 possible values for each missing RMDQ score using multiple linear regression models containing age, gender, treatment group, and a number of clinical outcomes as covariates.[14] The final imputed value was the arithmetic mean of the 5 data values created, with reported standard deviations and confidence intervals adjusted to account for the additional 'between imputation' variance. Baseline imbalances between treatment groups for RMDQ scores were controlled for using a multiple regression based adjustment.

Results

The sample sizes for the 4 analyses are described below:

(i) A PP analysis on the observed data, n=232 (58%). Eligibility for this analysis required individuals to have received their allocated treatment according to the study protocol, and provide RMDQ scores at baseline and 12 month follow-up.

(ii) An ITT analysis on the observed data, n=329 (82%). Individuals that provided RMDQ scores at baseline and 12 months are included in this analysis, irrespective of the treatment they received (i.e. includes protocol violators).

(iii) A PP analysis on the imputed data set, n=267 (66%). All individuals defined as receiving their treatment according to the study protocol are included in this analysis.

(iv) ITT analysis on the imputed data set, n=402 (100%). As we have imputed all missing 12 month RMDQ scores, every individual in the trial is included in this analysis.

Table 1 reports RMDQ scores for the four analyses described above. In each analysis, the confidence interval for the difference in RMDQ change scores lies

comfortably within the equivalence margin of ±2 points. After controlling for baseline differences in RMDQ scores, the per-protocol analysis on the observed data was the only analysis that failed to demonstrate equivalence (lower bound of the confidence interval = -2.01). Unsurprisingly, this uncertain finding emerged from the analysis with the smallest sample size.

The statistical power of our four analyses, given the sample sizes, were 50%, 76%, 61% and 86% respectively. Irrespective of the reduced statistical power of the analyses, we have demonstrated the equivalence of the two interventions, given an equivalence margin of ±2 points on the RMDQ. As our equivalence margin was based on the effect difference postulated in the superiority trial, the relatively high statistical power for our analyses are likely to be atypical for most retrospective equivalence analyses. Adopting a narrower equivalence margin would have resulted in reduced statistical power.

Discussion

Leaving the conclusion of pragmatic RCTs as 'not significantly better' is unsatisfactory. If funding restrictions and recruitment problems mean that trials continue to be powered on the basis of superiority rather than equivalence, one option is to ensure that the design of such trials allow for analysis in equivalence terms. They could thus contribute to systematic reviews and meta-analyses of equivalence trials. In practice, other than specifying the equivalence margin in advance, and ensuring that data required for a per-protocol analysis are collected, this would have little effect on the conduct or cost of clinical trials.

The retrospective equivalence analysis reported here highlights the potential to broaden the conclusions that can be drawn from superiority trials that show 'no difference' between interventions. Retrospective analysis, by definition, requires *a posteriori* consideration of the equivalence margin. A weakness of this analysis is that the equivalence margin was not determined in advance, although there is strong evidence within the low back pain literature that such a margin would be deemed appropriate.[13]

Perhaps the most important limitation is whether lower powered results from equivalence trials would be worth having. Our example has shown that reduced statistical power does not necessarily prevent studies from demonstrating equivalence, although this will not always be the case. One response to this limitation is that it would at least enable other equivalence trials to build on these results in meta-analysis.

Table 1: RMDQ scores at 12 month follow-up (four different analyses)

	BPM mean (sd)	PT mean (sd)	Mean difference in change scores (95% C.I.)[†]
(i) PP analysis of observed data			
Number analysed	98	134	
Baseline score	13.62 (4.6)	13.33 (4.7)	-
Absolute score at 12 months	5.16 (6.1)	4.37 (5.3)	-
12 month change score[‡]	8.70 (6.6)	8.77 (6.1)	-0.06 (-1.73 to 1.60)
'Controlled' change score*	8.41	8.98	-0.58 (-2.01 to 0.86)
(ii) ITT analysis of observed data			
Number analysed	164	165	
Baseline score	13.97 (4.8)	13.21 (4.8)	-
Absolute score at 12 months	5.17 (5.7)	4.41 (5.5)	-
12 month change score[‡]	8.80 (6.4)	8.79 (6.1)	0.00 (-1.35 to 1.36)
'Controlled' change score*	8.55	9.04	-0.50 (-1.67 to 0.67)
(iii) PP analysis of imputed data set			
Number analysed	113	154	
Baseline score	13.62 (4.6)	13.33 (4.7)	-
Absolute score at 12 months	5.00 (6.1)	4.51 (5.9)	-
12 month change score[‡]	8.62 (6.6)	8.82 (6.6)	-0.20 (-1.82 to 1.41)
'Controlled' change score*	8.50	8.90	-0.40 (-1.79 to 0.99)
(iv) ITT analysis of imputed data set			
Number analysed	201	201	
Baseline score	13.77 (4.8)	13.29 (4.9)	-
Absolute score at 12 months	5.06 (6.2)	4.56 (6.8)	-
12 month change score[‡]	8.71 (6.6)	8.73 (7.3)	-0.02 (-1.39 to 1.35)
'Controlled' change score*	8.55	8.89	-0.33 (-1.50 to 0.83)

[†] Difference in mean scores (BPM - PT).

[‡] Change in RMDQ score from baseline.

* Controlled for baseline differences in RMDQ scores. Reported values are the predicted scores from the multiple regression equations.

Economic Evaluation of the Keele Trial

In accordance with the recommendations in the literature, we conducted a full economic evaluation examining the costs and consequences of both interventions.[2,15] A summary of this analysis is given below. For the issues addressed in this paper, the key issue surrounds the results rather than the adopted methodology.

Methods - Data Collection

Utility was measured by the EuroQol EQ5D at baseline, 3 months and 12 months, with health state valuations elicited from a large representative sample of the UK population (the York A1 tariff).[16] These health utility values allowed us to estimate the number of quality adjusted life years (QALYs) patients experienced during the 12 month period post randomisation using area-under-the-curve analysis.[17] We assumed that utility changed linearly between consecutive data collection points. A separate standardised questionnaire was self-completed to collect health care resource use data for the 12 month period post randomisation. Specifically, this questionnaire collected data on back pain related resource use for hospital inpatient stays, outpatient consultations and visits to other health care practitioners (both within the UK National Health Service (NHS) and private practice). Consultations with NHS primary health care providers (e.g. general practitioners, practice nurse etc.), prescribed medications, and over-the-counter purchases were also collected.

Methods - Cost Estimation

To acknowledge important cost implications beyond those of the UK NHS, a health care perspective was adopted for the economic analysis, including both public sector and private sector health care resources. NHS primary health care resources were costed as national averages;[18] with inpatient and outpatient episodes costed using the 2002 NHS reference costs.[19] All costs were expressed in pounds sterling at 2001/02 prices. As the follow-up period was 12 months, no discounting was applied. Due to the paucity of high quality unit cost data for private health care consultations or procedures, private care was costed as the NHS equivalent. The British National Formulary (BNF 42) was used to cost prescribed medication.[20]

Methods - Statistical Analysis

Analysis was carried out according to the intention-to-treat principle. Multiple imputation was used to address the issue of missing EQ5D scores and resource use data.[14] Confidence intervals around differences in mean costs, EQ5D scores and

QALYs were generated using conventional parametric methods and bias-corrected and accelerated bootstrapping (1000 replications) in order to investigate whether the conventional parametric approach is robust to the level of skewness in the data. Baseline utility imbalances between treatment groups were controlled for using a multiple regression based adjustment.[21]

An incremental approach was used in the analysis, with differences in costs and health outcomes expressed using an incremental cost-utility ratio (ICUR), i.e. the incremental cost-per-QALY. Uncertainty was addressed using 5000 bootstrapped replications of mean cost and QALY differences, and plotting these pairs on a cost-utility plane, where incremental QALYs are represented along the *x*-axis and incremental costs along the *y*-axis.[22] A threshold analysis was performed for the ICUR, using a cost-utility acceptability curve to estimate the probability that the new intervention (BPM) is cost-effective at specific cost-per-QALY thresholds.[23,24] Sensitivity analysis considered a complete-case analysis and variation in the unit costs of private health care, although the results are not reported here.

Results

Mean health care costs, EQ5D scores and QALYs are presented in Table 2 for each treatment group. Although there were no statistically significant differences between groups, the point estimates suggest that PT is both more effective and more costly, with a cost-per-QALY ratio of £2362. Controlling for baseline utility increased the difference in mean QALYs between groups at 12 months, indicating a baseline imbalance that favoured the BPM group. Similar confidence intervals were obtained using parametric and bootstrap methods, indicating that the parametric intervals were robust to the level of skewness in the data (bootstrapped intervals not reported).

Figure 1 shows the cost-utility plane with 5000 bootstrapped replications of incremental cost-utility pairs comparing BPM to PT. The majority of replications are located in the southwest quadrant (86%), which demonstrates that on average patients randomised to BPM gained fewer QALYs, but at a lower mean health care cost compared with the PT group. The presence of bootstrapped pairs in all four quadrants of a cost-utility plane denotes that any ICUR should be interpreted with caution.[25] Acceptability curves provide a clear way of interpreting these results (see Figure 2). Each point on the curve represents the probability that BPM is cost-effective at specific cost-per-QALY thresholds. The higher the willingness-to-pay threshold, the less likely BPM will be cost-effective. Reading from the curve shows that if the cost-per-QALY threshold was a conservative £10,000 per QALY gained,

Table 2: Health outcomes over 12 months - Main analysis and complete case analysis

Outcomes (EQ5D & QALYs)	BPM mean (sd)	PT mean (sd)	Mean difference (95% C.I.)[a]
Main Analysis (*n* = 402)			
Baseline EQ5D	0.702 (0.3)	0.696 (0.3)	0.006 (-0.05 to 0.06)
3 month EQ5D	0.758 (0.4)	0.789 (0.3)	-0.031 (-0.10 to 0.04)
12 month EQ5D	0.770 (0.3)	0.785 (0.3)	-0.014 (-0.08 to 0.05)
QALYs over 12 months	0.756 (0.3)	0.776 (0.2)	-0.020 (-0.07 to 0.03)
QALYs over 12 months (controlled for baseline EQ5D)	0.755[b]	0.777[b]	-0.022 (-0.07 to 0.02) *p* = 0.35 (2 d.p.)
Total health care cost (£)	142.33 (261.3)	194.52 (445.6)	-61.79 (-146.76 to 23.18) *p* = 0.15 (2 d.p.)

[a] Difference = BPM - PT. C.I.s were generated using conventional parametric methods. There were no statistically significant differences between the groups for each of the analyses ($p \geq 0.05$).
[b] Values are predicted scores from the multiple regression when controlling for baseline imbalances.

the chance that BPM is cost-effective is only 17% (i.e. an 83% chance that PT is the preferred option).

Discussion

The estimated ICUR of £2362 lies comfortably within previous recommendations by the National Institute for Health and Clinical Excellence (NICE).[26] Following the quantification of uncertainty around the cost-per-QALY ratio, we concluded that PT is likely to be cost-effective. However, we went on to say that the absence of a clinically superior treatment package does raise the option of providing a brief pain management service in significantly fewer clinical sessions, which may be an attractive option to health care providers.

Implications for Trial-Based Economic Evaluations

In order to address issues of efficiency in the provision of health care, conventional economic evaluations take the form of cost-effectiveness analysis (CEA), where outcome measures are usually primary clinical outcomes, or cost-utility analysis (CUA), where outcomes are expressed as utilities. The main focus of these studies is

Figure 1: Incremental Cost-Utility Plane Comparing BPM to PT

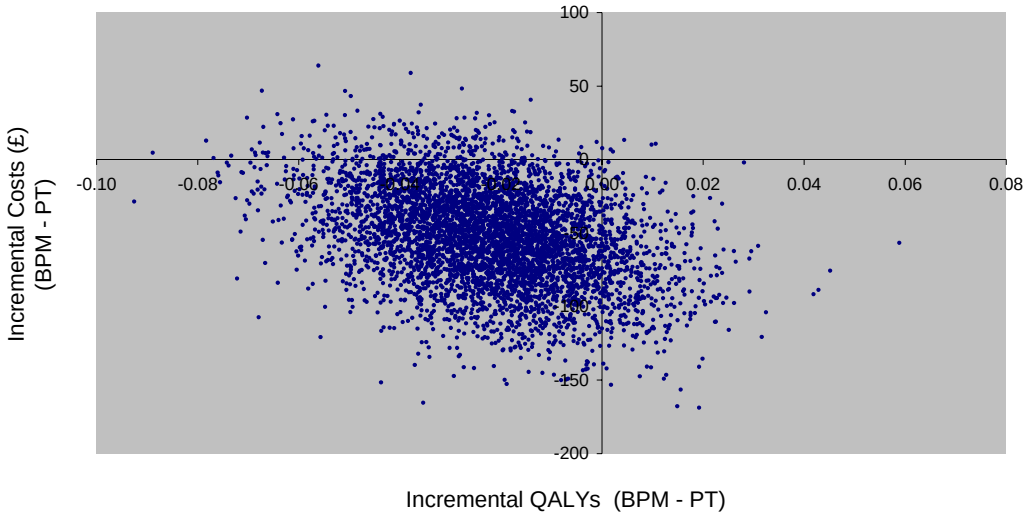
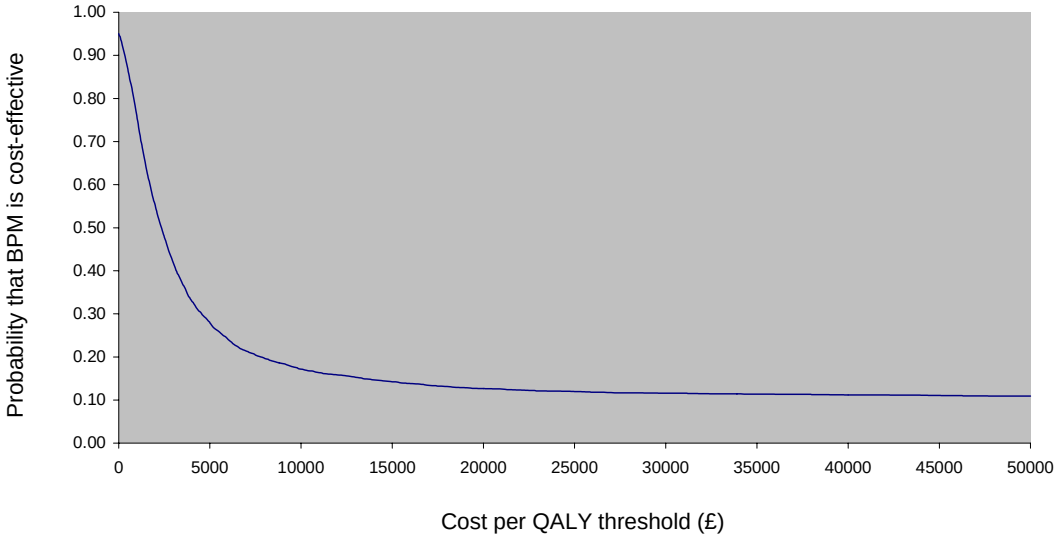


Figure 2: Cost-Utility Acceptability Curve Comparing BPM to PT



on the incremental cost-effectiveness ratio and the quantification of uncertainty around this ratio. An alternative form of analysis is a cost-minimisation analysis (CMA), where the researcher considers the effects of the interventions to be equal; consequently, only the costs of the interventions are compared. In a CMA, the primary interest is the cost difference between interventions, defined by the perspective of the study (e.g. societal, patient, or NHS perspective).[22] This may be

considered to be the appropriate form of analysis when an equivalence trial has demonstrated two or more interventions to be clinically indistinguishable.

Briggs and O'Brien have provided a comprehensive account of the circumstances under which CMA is an appropriate form of analysis, arguing that unless a study has been designed as an equivalence trial, the lack of a significant difference in effects does not provide a sufficient justification for CMA.[2] The main results from the economic evaluation of the low back pain trial are shown below (same results as in Table 2);

Outcome	BPM	PT	Difference (BPM -PT)	ICUR
Mean Health Care Cost	£142.33	£194.52	-£52.19	£2362
Mean QALYs	0.755	0.777	-0.022	

Following the quantification of uncertainty around the cost-per-QALY ratio, our cost-utility analysis concluded that physical therapy (PT) was likely to be the cost-effective management strategy for patients consulting primary care with low back pain of less than 12 weeks' duration. Earlier in this paper we reported the results of an equivalence analysis on the same trial data and demonstrated the equivalence of the two interventions, for an equivalence margin of ± 2 points on the RMDQ. Given the proposition by Briggs and O'Brien, and adopting a CMA design for the economic evaluation, our conclusions as to which treatment option would be deemed cost-effective would alter. We may conclude that there is evidence that the BPM package provides an equivalent clinical benefit to PT, at least cost. This result, coupled with the fact that BPM was delivered in significantly fewer treatment sessions, could be interpreted as evidence in favour of BPM being the cost-effective treatment option.

So…… which is the cost-effective treatment option?

<u>Discussion</u>

Performing equivalence analysis on data from superiority trials can provide a valuable addition to clinical trial analysis, particularly for pragmatic RCTs, adding to the possible conclusions that can be drawn. However, attention needs to be given to

the implications for trial-based economic evaluations. Using data from the Keele low back pain trial we have demonstrated that different conclusions can be drawn as to the cost-effectiveness of competing interventions, depending on the chosen method of analysis. Based on current recommendations in the health economic literature, both forms of analysis could arguably be justified.

This paper reports work-in-progress and at this stage we would like to use it to generate discussion on this topic. Three key issues we have considered are discussed below.

(i) What has been deemed 'equivalent'?

Equivalence may be assessed with regard to one primary outcome measure, typically a disease-specific measure of effectiveness (as in our low back pain example, also see Burgess *et al* [27]). Alternatively, equivalence may be assessed through a combination of outcome measures that cover a broader range of health benefits,[28] such as disease-specific measures, quality-of-life (QoL) measures and patients' own assessment of change in health status. The more rigorously researchers define equivalence, the less likely it becomes that the equivalence of competing interventions will be established.

In RCTs, primary outcome measures are used in sample size calculations, and are the main focus for the analysis. It is likely that an equivalence analysis on superiority trial data will also focus on the primary outcome measure. In our retrospective equivalence analysis of the low back pain trial we demonstrated the equivalence of the two interventions based on a back-pain specific disability measure (the RMDQ), but have we really shown the interventions to be sufficiently equivalent to warrant an economic evaluation based purely on the difference in costs? Allocating health care resources equitably and efficiently in a publicly funded health care system requires us to go beyond consideration of disease-specific outcome measures alone, as these *may* fail to capture improvement in quality of life or patients' perceived change in health status.

We propose that a narrow definition of equivalence, e.g. based on a disease-specific measure of health, is insufficient to justify the use of CMA. If, however, an equivalence trial (or equivalence analysis of a superiority trial) has demonstrated interventions to be indistinguishable across a range of outcome measures, covering different aspects of health, it would be appropriate to conduct CMA. This provides a greater methodological challenge, as equivalence margins would need to be defined for each outcome measure, reflecting the specific patient population within the trial. If

one considered *only* QALYs to matter, then equivalence in this health dimension would be sufficient. This would require establishing an equivalence margin for QALYs.

(ii) Economic evaluations with 'small' differences

Our proposition leaves open the possibility of performing cost-effectiveness or cost-utility analysis alongside RCTs that have shown no difference between interventions, even if an equivalence analysis has deemed the effect difference between interventions to be clinically unimportant. By definition, these economic evaluations will incorporate outcome measures that show small differences between interventions. In these situations it may be necessary to consider the plausibility of small differences in outcomes against the range of secondary outcome measures in the trial. If there were a degree of consistency within the outcome measures (i.e. point estimates show that the small differences favour one particular intervention), any conclusions derived through cost-effectiveness and/or cost-utility analysis would have greater credence.

(iii) Contribution to meta-analyses

As noted above, equivalence analysis requires specifying an equivalence margin and analysis per protocol rather than ITT. Generally, equivalence trials will either have lower power or will require more patients than superiority trials. Consequently, meta-analysis of similar trials is likely to be even more important for equivalence than for superiority. However, such meta-analysis requires different reporting than that conventionally defined by Cochrane. The recent CONSORT guideline for reporting equivalence trials is a useful step forward but requires prospective setting of the equivalence margin.[29] We suggest that where equivalence analysis can be done using a retrospectively set margin, this should be done as a contribution to the meta-analysis. The wider issue of whether sufficient evidence will accumulate by this route remains to be seen.

References:

1. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, Lam M, Seguin R. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med Res Methodol 2003; 3: 28.

2. Briggs AH, O'Brien BJ. The death of cost-minimisation analysis. Health Econ 2001;10(2):179-84.

3. Hay EM, Mullis R, Lewis M, Vohora K, Main CJ, Watson P, Dziedzic KS, Sim J, Minns Lowe C, Croft PR. Comparison of physical treatments versus a brief pain-management programme for back pain in primary care: a randomised clinical trial in physiotherapy practice. Lancet 2005; 365: 2024-30.

4. Whitehurst DGT, Lewis M, Yao G, Bryan S, Raftery JP, Mullis R, Hay E. A brief pain management programme and physical therapy for low back pain: results from an economic analysis alongside a randomised clinical trial. Arthritis Care & Research (in press).

5. Maniadakis N, Gray A. The economic burden of back pain in the UK. Pain 2000; 84(1): 95-103.

6. Royal College of General Practitioners. Guidelines for the management of acute low back pain. London: RCGP; 1996.

7. Clinical Standards Advisory Group on Back Pain. Back pain. London: HMSO Stationery Office; 1994.

8. Bronfort G, Haas M, Evans RL, Bouter LM. Efficacy of spinal manipulation and mobilization for low back pain and neck pain: a systematic review and best evidence synthesis. Spine J 2004; 4(3): 335-56.

9. Assendelft WJJ, Morton SC, Yu EI, Suttorp MJ, Shekelle PG. Spinal Manipulative Therapy for Low Back Pain. Ann Intern Med 2003; 138(11): 871-900.

10. Thomas E, Croft PR, Paterson SM, Dziedzic K, Hay EM. What influences participants' treatment preference and can it influence outcome? Results from a primary care-based randomised trial for shoulder pain. Br J Gen Pract 2004; 54 (499): 93-96.

11. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996; 313: 36-9.

12. Alderson P. Absence of evidence is not evidence of absence. BMJ 2004; 328:476-7.

13. Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference. Low back pain: outcome measures. J Rheum 2001; 28: 431-38.

14. Schafer JL. NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2. Software for Windows 95/98/NT, available from http://www.stat.psu.edu/~jls/misoftwa.html, 1999. (accessed 29 June 2006)

15. Altman DG, Bland MJ. Statistical notes: absence of evidence is not evidence of absence. BMJ 1995; 311: 485

16. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ 1996; 5(2): 141-54.

17. Matthews JNS, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. BMJ 1990; 300: 230-235.

18. Netten A, Dennett J, Knight J. Unit costs of health and social care. Canterbury: Personal Social Services Research Unit, University of Kent; 2002.

19. NHS Executive. National Schedule of Reference Costs, 2002. http://www.dh.gov.uk/PolicyAndGuidance/OrganisationPolicy/FinanceAndPlanning/NHSReferenceCosts/fs/en. (accessed 29 June 2006)

20. British Medical Association, Royal Pharmaceutical Society of Great Britain. British National Formulary. 42nd ed. London: BMJ Books; 2001.

21. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ 2005; 14(5): 487-96.23.

22. Korthals-de Bos I, van Tulder M, van Dieten H, Bouter L. Economic evaluations and randomized trials in spinal disorders: principles and methods. Spine 2004; 29(4): 442-8.

23. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making 1998; 18(2 Suppl): S68-80.

24. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Econ 2001; 10(8): 779-87.

25. Briggs A, Gray A. Handling uncertainty when performing economic evaluation of health care interventions. Health Technol Assess 1999; 3(2).

26. Raftery J. NICE: faster access to modern treatments? Analysis of guidance on health technologies. BMJ 2001; 323: 1300-3.

27. Burgess IF, Brown CM, Lee PN. Treatment of head louse infestation with 4% dimeticone lotion: randomised controlled equivalence trial. BMJ 2005; 330: 1423-7.

28. Tugwell PS, Wells GA, Shainhouse JZ. Equivalence study of a topical diclofenac solution (pennsaid) compared with oral diclofenac in symptomatic treatment of osteoarthritis of the knee: a randomized controlled trial. J Rheum 2004; 31:2002-12.

29. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW. Reporting of noninferiority and equivalence randomized trials: an extension to the CONSORT statement. JAMA 2006; 295(10): 1152-60.