BMJ

RESEARCH

Doctors' versus patients' global assessments of treatment effectiveness: empirical survey of diverse treatments in clinical trials

Evangelos Evangelou, research associate, Georgios Tsianos, research associate, John P A Ioannidis, professor

Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece

Correspondence to: J P A Ioannidis jioannid@cc.uoi.gr

doi:10.1136/bmj.39560.759572.BE

ABSTRACT

Objective To examine whether doctors' global assessments of treatment effects agree with patients' global assessments.

Design Survey of trials included in systematic reviews of treatments for diverse conditions.

Data sources Cochrane database of systematic reviews.

Data extracted Data on patients' global assessments and on doctors' global assessment for the same treatment against the same comparator.

Main outcome measures Relative odds ratio (ratio of odds ratios of global improvement with the experimental intervention versus control according to doctors compared with patients), and improvement rates according to doctors and patients.

Results Doctors' global assessments were compared with patients' global assessments for 63 different treatment comparisons (240 trials) in 18 conditions. The summary relative odds ratio across the comparisons was not significant (0.98, 95% confidence interval 0.88 to 1.08; I^2=0%, 95% confidence interval 0% to 30%). In 62 of the 63 comparisons the effects of treatment rated by patients and by doctors did not differ beyond chance, but for single comparisons the confidence intervals were large. Rates of improvement on average did not differ between doctors' assessments and patients' assessments (summary relative odds ratio 0.98, 0.88 to 1.06; I^2=0%, 0% to 24%).

Conclusion Doctors' global assessments of the effects of treatments are on average similar to those of patients.

INTRODUCTION

For several diseases and treatments the global assessment of change in disease status by patients and doctors are key outcomes for determining whether a treatment is effective. For some conditions other types of measurements besides an overall (global) impression are difficult, impractical, costly, or even non-existent. Global assessments have become popular choices as end points in selected disciplines, such as rheumatology, psychiatry, and dermatology, particularly when a single laboratory measurement or clinical measurement or documentation of an event cannot be used to adequately describe what happened to a study participant.

An important question is whether patients and doctors agree in their assessment of treatment outcomes. Self assessment by patients may avoid bias by an external assessor, whereas doctors may be more objective than their patients. Doctors may consider additional aspects of conditions that are not assessable by patients and may have insight into whether patients tend to amplify or minimise symptoms.[1] In theory, biases may be more likely when a study does not use blinding of doctors or patients, such as when blinding is impossible or compromised. Moreover, in different circumstances and for different diseases biases may operate differently between patients and doctors—some patients with mental or neurological diseases, for example, may be biased or inaccurate in the appraisal of their condition. Similarly, doctors may be inaccurate when they have few or no objective signs and tests on which to base their observations and have to use primarily patient reported information.

Several studies have evaluated whether global assessment in specific conditions and settings is more appropriately done by patients than by doctors. Some studies suggest that patients' opinions do not agree with those of doctors even though they are measuring the same outcome.[2-4] Other studies, however, showed little difference between self reported assessment and doctors' assessment.[5 6] Evidence is lacking as to whether differences in appraisals also result in systematic differences in the estimates of treatment effects in clinical trials. For example, a meta-analysis of trials on the interleukin 1 receptor antagonist in rheumatoid arthritis suggested that patient reported outcomes provided more favourable estimates of treatment effects than outcomes reported by doctors.[7]

We obtained empirical information on the possible extent of discordance between doctors' and patients' global assessments of treatment effects in clinical trials for various diseases and treatments. We evaluated a sample of systematic reviews of clinical trials where both patients' and doctors' impressions of global

improvement had been used as outcomes to evaluate the same treatment.

METHODS

We considered published systematic reviews from the Cochrane Library (Issue 3, 2006) that included separate quantitative analyses (meta-analyses) of doctors' and patients' global assessment at the same time point for the comparison of the same experimental treatment against the same comparator (placebo, no treatment, or other treatment). We accepted comparisons regardless of the number of trials with data for each type of assessment outcome and regardless of whether such studies were the same, overlapping, or different. We excluded protocols and reviews that had been withdrawn. We also excluded comparisons where we could not clearly define the experimental treatment between two active comparators. Whenever global assessment was done at several different time points we retained the data for the time point where the largest number of studies would have available data for either type of assessment outcome. We accepted reviews regardless of whether the global assessment pertained to change in binary outcomes (improvement, deterioration, cure, failure, success) or to change in scores for continuous outcomes.

We searched the Cochrane Library database using the term "global". We also searched a random sample of 200 Cochrane reviews using the terms "patient assessment" or "clinician assessment" to check that we had not missed possible eligible reviews that did not use the term "global". The retrieved reviews were screened for eligibility, first by examining the tables and figures and, if in doubt, by examining the full text. Eligible reviews could contain more than one comparison with different treatments or comparators. For example, within a review we might assess the global effectiveness of a treatment compared with standard treatment and assess the global effectiveness of the same treatment compared with placebo. We counted and evaluated eligible comparisons. Finally, we searched all Cochrane systematic reviews on diseases where at least three eligible comparisons had already been identified through the search strategy.

In each eligible comparison we recorded the studies that had data on doctors' global assessments and those that had data on patients' global assessments and noted any overlap. For each of these studies we recorded the year of publication, first author, outcome definition for global change, and the 2×2 tables or the mean difference and standard deviation per arm for global change according to both the doctors and the patients.

Binary and continuous outcomes

We calculated the odds ratio of both doctors' and patients' assessments and the variances of their natural logarithms. We consistently coined the comparisons to reflect the contrast of the experimental treatment with comparator (placebo, no treatment, other treatment) and consistently to reflect improvement rather than deterioration. This means that when the data reflected the number of patients who deteriorated (for example, 12/30), we took the complementary counts (that is, 18/30); whenever the experimental treatment was better, this was coined to be consistently an odds ratio greater than 1.

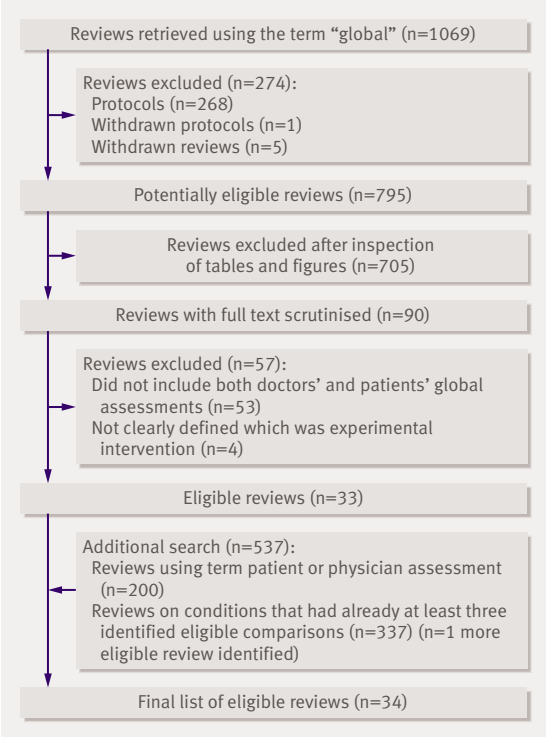

Fig 1 | **Flow chart of selected reviews**

We calculated the weighted standardised mean differences of the continuous outcomes and transformed them to odds ratios[8] using a formula that incorporates the Hedges' g, a measure that quantifies continuous outcomes using standardised mean differences.[9] All comparisons were consistently coined as for the binary outcomes.

Analyses

For each comparison we combined the natural logarithms of the odds ratio of both doctors' and patients' assessments across each of the eligible studies to obtain the summary effect of the odds ratio of assessments for doctors and for patients. Then we compared the ratio of the summary odds ratio of doctors' assessments with the summary odds ratio of patients' assessments to obtain the relative odds ratio for each comparison. A relative odds ratio exceeding 1 equates to the doctors' assessments giving a more favourable response for the experimental treatment than the patients' assessments. A relative odds ratio less than 1 equates to the doctors' assessments giving a less favourable response for the experimental treatment than the patients' assessments. The variance of the natural logarithm of the relative odds ratio is the sum of the variances of the natural logarithms of the odds ratio of the doctors' assessments and the odds ratio of the patients' assessments.

We combined the estimates of the natural logarithm of the relative odds ratio across all comparisons to obtain the summary natural logarithm of relative odds ratio,[10 11] using fixed effects and random effects.[12 13] We used the Cochran's Q statistic (considered statistically significant for $P<0.10$) and the I^2 metric to quantify heterogeneity between comparisons in the estimates of the natural logarithm of the relative odds ratio.[14] I^2 is independent of the number of comparisons and a value of 50% or more reflects sizeable heterogeneity. We also provide 95% confidence intervals for I^2 in the main analyses.[14 15] In the absence of heterogeneity ($I^2=0$), random and fixed effects coincide.

For the main analysis we considered all eligible comparisons. We also carried out sensitivity analyses, limited to comparisons when all studies had both doctors' and patients' assessments or to trials that had both doctors' and patients' assessments. In these situations outcomes are directly paired, so we estimated a natural logarithm of the relative odds ratio for each study before combining these to obtain a summary value.

Furthermore, we carried out subgroup analyses according to condition, with the conditions merged into three categories: musculoskeletal, neuropsychiatric and pychosomatic, and other. Additional subgroup analyses were done according to type of assessment outcome (binary or continuous); whether both doctors and patients were blinded, only doctors were blinded, only patients were blinded, or neither were blinded; and whether the comparison referred to treatment compared with no treatment or placebo or to two active treatments.

Finally, doctors' and patients' assessments may agree at the level of the relative treatment effect (odds

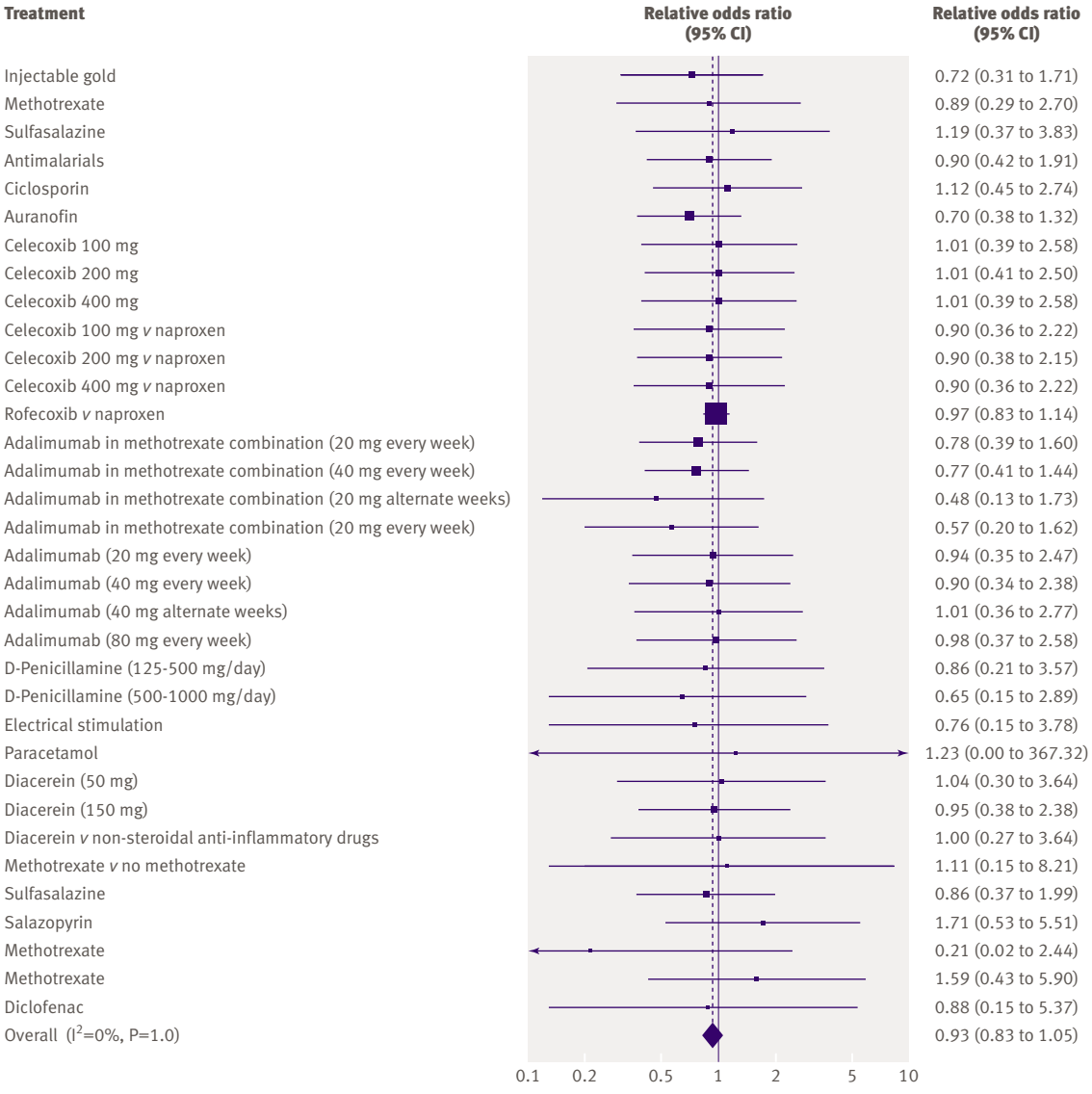

Fig 2 | **Relative odds ratios (95% confidence intervals) for patients' compared with doctors' assessments on effectiveness of treatments for musculoskeletal conditions. Experimental treatment is compared with standard treatment, older treatment, or no treatment (or placebo). Diamond shows summary relative odds ratio according to random effects calculations. For details on the disease or condition and Cochrane review for each comparison see supplementary table at www.dhe.med.uoi.gr/sup_mat.php/**

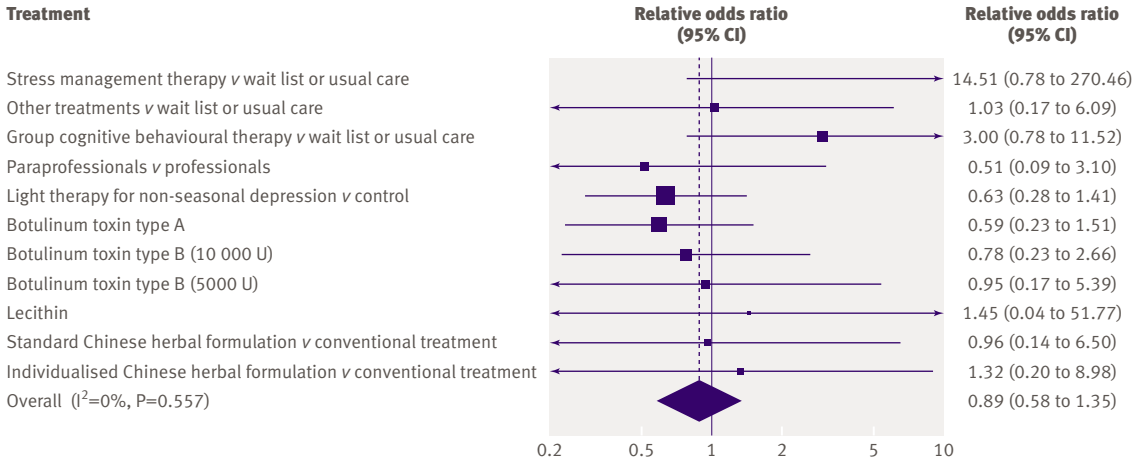

Fig 3 | **Relative odds ratios (95% confidence intervals) for patients' compared with doctors' assessments on effectiveness of treatments for neuropsychiatric and psychosomatic conditions. Experimental treatment is compared with standard treatment, older treatment, or no treatment (or placebo). Diamond shows summary relative odds ratio according to random effects calculations. For details on the disease or condition and Cochrane review for each comparison see supplementary table at www.dhe.med.uoi.gr/sup_mat.php/**

ratio) but may disagree on the absolute proportion of patients who improve in both arms. Therefore we also examined whether the overall proportions showing improvement differed between doctors and patients. We limited these analyses to the set of studies where data on both doctors' and patients' assessments were available for the same study. For these evaluations we combined both arms (experimental and control) for each type of outcome. For binary outcomes we estimated the total number of patients who had improved among the total of patients in the experimental and control arms combined. For continuous outcomes we estimated a common mean effect and variance, combining the respective measures of the experimental and control arms by fixed effects. Then we estimated the odds ratio of global improvement according to doctors and according to patients. For continuous outcomes we used the Hedges g transformation. We combined the estimates for the natural logarithm of the odds ratio for improvement across studies for each comparison. These summary estimates were then combined across comparisons. This was done in a similar fashion to the natural logarithm of the relative odds ratio.

All analyses were done in Intercooled STATA 8.2. P values are two tailed.

RESULTS

Figure 1 shows the flow of the reviews. Thirty four reviews[w1-w34] totalling 63 comparisons (n=240 studies) were eligible for analysis (see details of comparisons at www.dhe.med.uoi.gr/sup_mat.php/). A variety of conditions and treatments were evaluated, with 34 comparisons of musculoskeletal conditions (rheumatoid arthritis, osteoarthritis, elbow pain, psoriatic arthritis, juvenile arthritis, ankylosing spondylitis),[w1-w18] 11 comparisons of neuropsychiatric or psychosomatic conditions (post-traumatic stress disorder, anxiety, depression, alcohol withdrawal, tardive dyskinesia, cervical dystonia, irritable bowel syndrome),[w19-w25] and 18 comparisons of other conditions (asthma, acne, surgical incision, skin photodamage, rosacea, prostatic hyperplasia).[w26-w34]

In 44 comparisons (118 studies) perfect overlap of studies occurred (the same studies had data on doctors' and patients' assessment), in 17 comparisons (115 studies) partial overlap occurred, and in two comparisons (7 studies) no overlap occurred. Thirty two comparisons referred to continuous outcomes (perfect overlap n=25, partial overlap n=5, no overlap n=2) and 31 comparisons referred to binary outcomes (perfect overlap n=19, partial overlap n=12; see www.dhe.med.uoi.gr/sup_mat.php/).

Data synthesis

The summary results across the 63 comparisons showed overall agreement for the global estimate of treatment effectiveness between doctors and patients. The summary relative odds ratio was not significant (0.98, 95% confidence interval 0.88 to 1.08) and no significant heterogeneity was observed across the comparisons (I^2=0%, 95% confidence interval 0% to 30%; Cochran's Q P=0.99). Treatment effects according to patients and doctors did not differ beyond chance for 62 of the 63 comparisons, whereas for long acting β_2 agonists in asthma doctors gave a significantly more favourable appraisal of effectiveness than did patients (relative odds ratio 2.86, 1.48 to 5.55). Most point estimates of relative odds ratios for specific comparisons were close to 1. On the basis of point estimates, the most unfavourable relative perception of doctors' global assessment was in the use of methotrexate to treat psoriatic arthritis (relative odds ratio 0.21, 0.02 to 2.44)[w16] whereas the most favourable was for the implementation of stress management therapy for post-traumatic stress disorder (relative odds ratio 14, 0.78 to 270).[w19]

When the analysis was restricted to the 44 comparisons (n=118 studies) with perfect overlap of studies the results were practically identical. The summary relative odds ratio showed no difference between doctors and patients (0.97, 0.87 to 1.09; I^2=0%, P for heterogeneity 1.00). For the 17 comparisons with partial overlap (115 studies), data from doctors and patients were available in only some of the trials (n=76). When the analysis concerned the 194 trials that had data from doctors and patients (61 comparisons), the summary relative odds ratio was not significant (0.96, 0.86 to 1.07; I^2=0%, P for heterogeneity 0.99).

Subgroup analyses

Despite some trends for more favourable appraisal by patients of effectiveness in musculoskeletal conditions (fig 2) and neuropsychiatric or psychosomatic conditions (fig 3) and by doctors in other conditions (fig 4), the observed differences were not beyond chance (table). The estimated treatment effects did not differ depending on type of outcome (continuous *v* binary) or type of comparator.

In most comparisons (52/63) both patients and doctors were reported to be blinded. In these comparisons no evidence was found of a difference between doctors and patients (relative odds ratio 0.94, 95% confidence interval 0.85 to 1.04). In six comparisons (post-traumatic stress disorder, light therapy for non-seasonal depression, and closure of surgical incision) only the doctor was blinded; the relative odds ratio was 1.81 (0.79 to 4.16), but considerable heterogeneity existed between studies (I^2=49%). The blinded doctors tended to give more favourable assessments for the effectiveness of experimental treatments for post-traumatic stress disorder and closure of surgical incisions than did the patients, but the opposite trend was seen for light therapy for non-seasonal depression. In four comparisons no adequate information was provided on blinding: relative odds ratio 1.27 (0.73 to 2.22). In one comparison, blinding of patients was not possible and it was not stated whether the doctors were blinded (psychological treatment for anxiety and depression by paraprofessionals *v* professionals); the relative odds ratio showed a non-significant trend for more favourable appraisal of effectiveness by patients.

Rates of improvement

Rates of improvement did not differ between doctors' and patients' assessments (summary relative odds ratio 0.98, 95% confidence interval 0.88 to 1.06; I^2=0%, 0% to 24%). This meant that for an improvement rate of 10% according to patients the expected average improvement rate according to doctors would be 9.8% (8.9% to 10.5%) and that for an improvement rate of 40% according to patients the expected average improvement rate according to doctors would be 39.5% (37.0% to 41.4%).

The random effects summary relative odds ratio for improvement for musculoskeletal conditions was 0.95 (0.84 to 1.06, I^2=0%), for neuropsychiatric or psychosomatic conditions was 0.91 (0.60 to 1.33, I^2=0), and for other conditions was 1.06 (0.89 to 1.22, I^2=3%).

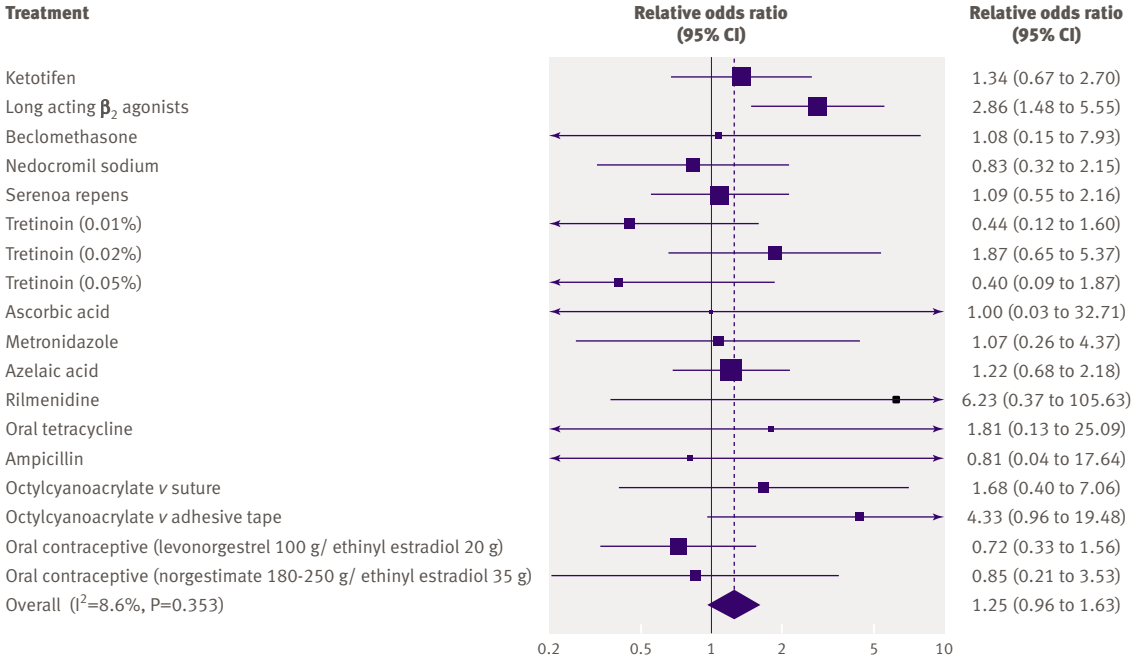

Fig 4 | Relative odds ratios (95% confidence intervals) for patients' compared with doctors' assessments on effectiveness of treatments for other conditions. Experimental treatment is compared with standard treatment, older treatment, or no treatment (or placebo). Diamond shows summary relative odds ratio according to random effects calculations. For details on the disease or condition and Cochrane review for each comparison see supplementary table at www.dhe.med.uoi.gr/sup_mat.php/

DISCUSSION

In this empirical evaluation we found on average an overall agreement between patients' and doctors' global assessments of effectiveness for diverse treatments. We detected no notable heterogeneity across the evaluated treatments, but the uncertainty in the results for single comparisons was typically large. Thus we cannot exclude the possibility of modest differences between specific treatments in particular diseases and settings. Furthermore, on average the rates of improvement were similar according to the appraisal of patients and doctors.

Most clinical questions have limited evidence from clinical trials and thus the uncertainty in the estimated treatment effects is often large, when only one topic is examined. By examining a large number of comparisons a more precise average emerges.

The previous literature on patients' and doctors' appraisals of outcome has dealt mostly with musculoskeletal diseases, along with other conditions such as cancer and asthma.[1-6 16-21] Several studies have focused on the considerable discrepancies between these assessments. For example, patients with cancer rate their health status differently from their doctors, and different doctors can give different ratings for the same patient.[20] Doctors may underestimate the needs of patients[21] or fail to recognise functional disability.[18] Surveys in musculoskeletal diseases have shown that patients and doctors often focus on different aspects of the disease: doctors prefer objective clinical signs or tests whereas patients focus more on their psychological wellbeing.[3 4 17] It is impossible to say in each study and case how much patients and doctors focused on wellbeing or on disease activity. Different patients and doctors may have different perspectives. Differences may average out on large samples and the estimated treatment effects may remain unaffected. Nevertheless, differences between patients' and doctors' assessments may still be important for the management of individual patients or for making a correct diagnosis (for example, patients with rheumatoid arthritis *v* patients with fibromyalgia).[22]

Most of the comparisons we analysed were in trials where all assessors of outcome were blinded. In theory, if blinding is not violated then patients and doctors should not be biased in appraising the effectiveness of a treatment. Our results are consistent with this interpretation. The more limited data on circumstances in which blinding was not achieved show non-significant deviations between patients' assessments and those of doctors. Nevertheless, for trials where only patients were unblinded we observed mostly trends for less favourable estimates of effectiveness by patients (table). Thus bias due to lack of blinding was unlikely to lead to more optimistic results.

For many comparisons we found no full overlap of the studies. Therefore we carried out sensitivity analyses only when studies were fully matched. The results were almost identical. We did not, however, have individual level data to examine whether the same or different patients were thought to improve according to patients and doctors.

Finally, concordance between patients' and doctors' assessments may be better in clinical trials than in everyday practice. The experimental nature of clinical trials may compel doctors to be more careful, meticulous, and comprehensive in assessing patient outcomes, and patients enrolled in clinical trials may be self selected. In all, the average agreement between patients and doctors in our empirical evaluation should not necessarily be interpreted as evidence that one of the two is redundant. For some conditions, such as

Summary relative odds ratios in various subgroups

Subgroup	Comparisons	Summary relative odds ratio (95% CI)		I^2 (95% CI)
		Fixed effects	Random effects	
Disease or condition:				
Musculoskeletal	34	0.93 (0.83 to 1.05)	0.93 (0.83 to 1.05)	0 (0 to 39)
Neuropsychiatric or psychosomatic	11	0.89 (0.58 to 1.35)	0.89 (0.58 to 1.35)	0 (0 to 60)
Other	18	1.26 (0.99 to 1.61)	1.25 (0.96 to 1.63)	9 (0 to 44)
Type of outcome:				
Continuous	32	0.94 (0.83 to 1.06)	0.94 (0.83 to 1.06)	0 (0 to 40)
Binary	31	1.08 (0.89 to 1.31)	1.08 (0.89 to 1.31)	0 (0 to 40)
Blinding:				
Patient and doctor blinded	52	0.96 (0.86 to 1.07)	0.96 (0.86 to 1.07)	0 (0 to 32)
Inadequate reporting of blinding	4	1.27 (0.73 to 2.22)	1.27 (0.73 to 2.22)	0 (0 to 85)
Only doctor blinded	6	1.38 (0.81 to 2.36)	1.81 (0.79 to 4.16)	49 (0 to 80)
Patient and doctor unblinded	1	0.51 (0.09 to 3.10)	0.51 (0.09 to 3.10)	NA
Comparisons:				
Treatment *v* placebo or no treatment	44	1.01 (0.86 to 1.18)	1.01 (0.86 to 1.18)	0 (0 to 35)
Two active treatments	19	0.96 (0.83 to 1.09)	0.96 (0.83 to 1.09)	0 (0 to 49)

NA=not applicable.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Global assessments by patients and doctors are commonly used to assess the effectiveness of treatments for various diseases

Some evidence suggests that assessments by patients may differ from those by doctors

WHAT THIS STUDY ADDS

Doctors' and patients' global assessments agreed on average on the derived estimates of treatment effects

Modest differences in either direction for specific conditions and treatments cannot be excluded

rheumatoid arthritis, both patients' and doctors' global assessments are typically used already.[16 23 24] In other diseases and trials when only one of the two types of assessment is used, consideration should be given to evaluating both and studying their relative performance in measuring treatment effects. The views of both patients and doctors may offer complementary information in clinical trials and in everyday practice.

We thank Peter Tugwell and Theodore Pincus for useful comments on the manuscript.

Contributors: JPAI had the original idea for this project and proposed the design. All authors worked on the protocol. EE and GT extracted the data and JPAI oversaw the collected data and arbitrated on discrepancies. EE did the statistical analysis with help from JPAI. All authors interpreted the data. JPAI and EE wrote the manuscript and all authors revised drafts and approved the final version. JPAI is guarantor.

Competing interests: None declared.

Ethical approval: Not required.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Bellamy N. Science of assessment. *Ann Rheum Dis* 2005;64(suppl 2):ii42-5.
2 Berkanovic E, Hurwicz ML, Lachenbruch PA. Concordant and discrepant views of patients' physical functioning. *Arthritis Care Res* 1995;8:94-101.
3 Kwoh CK, O'Connor GT, Regan-Smith MG, Olmstead EM, Brown LA, Burnett JB, et al. Concordance between clinician and patient assessment of physical and mental health status. *J Rheumatol* 1992;19:1031-7.
4 Neville C, Clarke AE, Joseph L, Belisle P, Ferland D, Fortin PR. Learning from discordance in patient and physician global assessments of systemic lupus erythematosus disease activity. *J Rheumatol* 2000;27:675-9.
5 Hidding A, van Santen M, De Klerk E, Gielen X, Boers M, Geenen R, et al. Comparison between self-report measures and clinical observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. *J Rheumatol* 1994;21:818-23.
6 Jacobs JW, Oosterveld FG, Deuxbouts N, Rasker JJ, Taal E, Dequeker J, et al. Opinions of patients with rheumatoid arthritis about their own functional capacity: how valid is it? *Ann Rheum Dis* 1992;51:765-8.
7 Cohen SB, Strand V, Aguilar D, Ofman JJ. Patient- versus physician-reported outcomes in rheumatoid arthritis patients treated with recombinant interleukin-1 receptor antagonist (anakinra) therapy. *Rheumatology (Oxford)* 2004;43:704-11.
8 Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychol Bull* 1995;117:167-78.
9 Cooper H, Hedges L, eds. *The handbook of research synthesis*. New York: Russel Sage Foundation, 1994.
10 Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287:2973-82.
11 Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;21:1513-24.
12 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
13 Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997;127:820-6.
14 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58.
15 Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. *BMJ* 2007;335:914-6.
16 Pincus T, Strand V, Koch G, Amara I, Crawford B, Wolfe F, et al. An index of the three core data set patient questionnaire measures distinguishes efficacy of active treatment from that of placebo as effectively as the American College of Rheumatology 20% response criteria (ACR20) or the Disease Activity Score (DAS) in a rheumatoid arthritis clinical trial. *Arthritis Rheum* 2003;48:625-30.
17 Yen JC, Abrahamowicz M, Dobkin PL, Clarke AE, Battista RN, Fortin PR. Determinants of discordance between patients and physicians in their assessment of lupus disease activity. *J Rheumatol* 2003;30:1967-76.
18 Calkins DR, Rubenstein LV, Cleary PD, Davies AR, Jette AM, Fink A, et al. Failure of physicians to recognize functional disability in ambulatory patients. *Ann Intern Med* 1991;114:451-4.
19 Carr AJ, Donovan JL. Why doctors and patients disagree. *Br J Rheumatol* 1998;37:1-4.
20 Slevin ML, Plant H, Lynch D, Drinkwater J, Gregory WM. Who should measure quality of life, the doctor or the patient? *Br J Cancer* 1988;57:109-12.
21 Canonica GW, Baena-Cagnani CE, Blaiss MS, Dahl R, Kaliner MA, Valovirta EJ, et al. Unmet needs in asthma: global asthma physician and patient (GAPP) survey: global adult findings. *Allergy* 2007;62:668-74.
22 DeWalt DA, Reed GW, Pincus T. Further clues to recognition of patients with fibromyalgia from a simple 2-page patient multidimensional health assessment questionnaire (MDHAQ). *Clin Exp Rheumatol* 2004 Jul-Aug;22(4):453-61.
23 Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomized controlled trial. *Arthritis Rheum* 1995;38:1568-80.
24 American College of Rheumatology Committee to Reevaluate Improvement Criteria. A proposed revision to the ACR20: the hybrid measure of American College of Rheumatology response. *Arthritis Rheum* 2007;57:193-202.

Accepted: 7 April 2008