

### WHAT IS ALREADY KNOWN ON THIS TOPIC

Fetal scalp blood can be tested for hypoxia in fetuses with worrying intrapartum heart rate traces

pH analysis has a sampling or analysis failure rate of 11-20% and has been excluded from clinical practice in the US

### WHAT THIS STUDY ADDS

There was no difference between assessment of fetal acidaemia with lactate analysis or pH analysis

At worst, replacement of pH measurement with lactate measurement could result in a one third increase in acidaemia at delivery

Apgar scores at five minutes, or admissions to neonatal intensive care units. Sampling failure was more common in the pH group. Combined analyses are not recommended as they are likely to increase the number of interventions without decreasing metabolic acidaemia at birth.

We thank all the women who took part in the trial and the obstetricians and midwives who recruited patients and managed their labours. The following hospitals took part: Danderyd, Kalmar, Karlstad, Karolinska, Lund, Söder Hospital, Trollhättan, Örebro, Gothenburg, Linköping. We also thank Hans Pettersson, Department of Biostatistics, Karolinska Institute at Söder Hospital, Stockholm, Sweden, for excellent statistical advice. The independent steering committee comprised Sven Cnattingius, Bengt Persson, Magnus Westgren, Karolinska Institutet, Stockholm, Sweden.

**Contributors:** See bmj.com.

**Funding:** Signhild Engqvists Stiftelse, Allmänna BB's Minnesfond, the regional city council research and development foundations, the health and medical committee of the region Västra Götaland, and Medexa, Lomma, Sweden.

**Competing interests:** EW-I and LN are shareholders in Obstecare, a company dealing with development of IT based decision support for labour dystocia. No product is yet on the market.

**Ethical approval:** Karolinska Institutet, Stockholm, Sweden (file record 109/02).

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

- 1 Bretscher J, Saling E. pH values in the human fetus during labor. *Am J Obstet Gynecol* 1967;97:906-11.
- 2 Westgren M, Kuger K, Ek S, Grunevald C, Kublickas M, Naka K, et al. Lactate compared with pH analysis at fetal scalp blood sampling: a prospective randomised study. *Br J Obstet Gynaecol* 1998;105:29-33.
- 3 Tuffnell D, Haw WL, Wilkinson K. How long does a fetal scalp blood sample take? *BJOG* 2006;113:332-4.
- 4 Schimojo N, Naka K, Uenoyama H, Hamamoto K, Yoshioka K, Okuda K. Electrochemical assay system with single-use electrode strip for measuring lactate in whole blood. *Clin Chem* 1993;39:2312-4.
- 5 Samat HB, Samat MS. Neonatal encephalopathy following fetal distress: a clinical and electroencephalographic study. *Arch Neurol* 1976;33:696-705.
- 6 Goldaber KG, Gilstrap LC, Leveno KJ, Dags JS, McIntire DD. Pathologic fetal acidemia. *Obstet Gynecol* 1991;78:1103-7.
- 7 Low JA, Panagiotopoulos C, Derrick EJ. Newborn complications after intrapartum asphyxia with metabolic acidosis in the term fetus. *Am J Obstet Gynecol* 1994;170:1081-7.
- 8 MacLennan A. A template for defining a causal relation between acute intrapartum events and cerebral palsy: international consensus statement. *BMJ* 1999;319:1054-9.
- 9 Kruger K, Hallberg B, Blennow M, Kublickas M, Westgren M. Predictive value of fetal scalp blood lactate concentration and pH as marker for neurologic disability. *Am J Obstet Gynecol* 1999;181:1072-8.
- 10 Clark SL, Paul RH. Intrapartum fetal surveillance: the role of fetal scalp blood sampling. *Am J Obstet Gynecol* 1985;153:717-20.
- 11 Nordström L, Ingemarsson I, Kublickas M, Persson B, Shimojo N, Westgren M. Scalp blood lactate—a new test strip method for monitoring fetal well-being in labour. *Br J Obstet Gynaecol* 1995;102:894-9.

Accepted: 31 March 2008

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

Evangelos Evangelou, Georgios Tsianos, John P A Ioannidis

### EDITORIAL by Spertus

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

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

BMJ 2008;336:1287-90  
doi:10.1136/bmj.39560.759572.BE

This article is an abridged version of a paper published on bmj.com. Cite this article as: BMJ 2008, 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

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<sup>1-3</sup> whereas others show little difference.<sup>4,5</sup>

We obtained empirical information on the extent of discordance between doctors' and patients' global assessments of treatment effects in clinical trials for various diseases and treatments.

## METHODS

We considered systematic reviews from the Cochrane Library that included separate quantitative 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 whether such studies were the same, overlapping, or different. When global assessment was done at several time points we retained the data for the time where the largest number of studies would have data for either type of assessment outcome.

Eligible reviews could contain more than one comparison with different treatments or comparators. In each 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 outcome definition for global change and the 2×2 tables or the mean difference and standard deviation per arm for global change according to the doctors and patients.

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 and consistently to reflect improvement rather than deterioration. This means that when the data reflected the number of patients who had deteriorated, we took the complementary counts; whenever the experimental treatment was better, this was coined to be consistently an odds ratio greater than 1.

We calculated the weighted standardised mean differences of the continuous outcomes and transformed them to odds ratios<sup>6</sup> using a formula that incorporates the Hedges' *g*.<sup>7</sup> 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 that of patients' assessments to obtain the relative odds ratio. A relative odds ratio exceeding 1 equates to the doctors' assessments giving a more favourable response for the experimental treatment than the patients' assessments.

We combined the estimates of the natural logarithm of the relative odds ratio across comparisons to obtain the summary natural logarithm of relative odds ratio,<sup>8</sup> using fixed effects and random effects.<sup>9</sup> We used the Cochran's *Q* and *I*<sup>2</sup> statistics to quantify heterogeneity between comparisons in the estimates of the natural logarithm of the relative odds ratio.<sup>10</sup>

We carried out sensitivity analyses, limited to comparisons when all studies had both doctors' and patients' assessments or to trials that had 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 for musculoskeletal, neuropsychiatric and psychosomatic, and other conditions. Additional subgroup analyses were done according to type of assessment outcome; whether doctors and patients were blinded, doctors were blinded, 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.

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 doctors' and patients' assessments were available for the same study. For these evaluations we combined both arms for each type of outcome. For binary outcomes we estimated the number of patients who had improved among those 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 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. *P* values are two tailed.

## RESULTS

Thirty four reviews<sup>w1-w34</sup> totalling 63 comparisons (*n*=240 studies) were eligible for analysis (see [bmj.com](http://bmj.com) and [www.dhe.med.uoi.gr/sup\\_mat.php/](http://www.dhe.med.uoi.gr/sup_mat.php/)). Several conditions and treatments were evaluated, with 34 comparisons of musculoskeletal conditions,<sup>w1-w18</sup> 11 comparisons of neuropsychiatric or psychosomatic conditions,<sup>w19-w25</sup> and 18 comparisons of other conditions.<sup>w26-w34</sup>

In 44 comparisons (118 studies) perfect overlap of studies occurred, in 17 comparisons (115 studies) partial overlap occurred, and in two comparisons (7 studies) no overlap occurred. Thirty two comparisons referred to continuous outcomes and 31 to binary outcomes (see [www.dhe.med.uoi.gr/sup\\_mat.php/](http://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 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  $\beta_2$  agonists in asthma doctors gave a significantly more favourable appraisal than 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: 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)<sup>w16</sup> 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).<sup>w19</sup>

When the analysis was restricted to the 44 comparisons with perfect overlap 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, data from doctors and patients were available in only some of the trials ( $n=76$ ). When the analysis concerned the 194 trials with 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 and neuropsychiatric or psychosomatic conditions and by doctors in other conditions (see [bmj.com](#)), the observed differences were not beyond chance (see [bmj.com](#)). The estimated treatment effects did not differ depending on outcome or comparator.

In most comparisons (52/63) patients and doctors were reported to be blinded. In these comparisons no difference was found between doctors and patients (relative odds ratio 0.94, 95% confidence interval 0.85 to 1.04). In six comparisons only the doctor was blinded (relative odds ratio 1.81, 0.79 to 4.16), but considerable heterogeneity existed between studies ( $I^2=49\%$ ). The doctors tended to give more favourable assessments for the effectiveness of the experimental treatments for post-traumatic stress disorder and closure of surgical incisions than 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; 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%).

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,  $I^2=3\%$ ).

### 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 treatments, but the uncertainty in the results for single comparisons was 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 patients and doctors.

The previous literature on patients' and doctors' appraisals of outcome has dealt mostly with musculoskeletal diseases.<sup>2-5 11-14</sup> Several studies have focused on the discrepancies between these assessments. Doctors may underestimate the needs of patients<sup>15</sup> or fail to recognise functional disability.<sup>13</sup> 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.<sup>2,3,12</sup> It is impossible to say in each study and case how much patients and doctors focused on wellbeing or on disease activity. 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.<sup>16</sup>

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.

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 meticulous 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

## 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

empirical evaluation should not necessarily be interpreted as evidence that one of the two is redundant.

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

**Contributors:** See bmj.com.

**Competing interests:** None declared.

**Ethical approval:** Not required.

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

- 1 Berkanovic E, Hurwicz ML, Lachenbruch PA. Concordant and discrepant views of patients' physical functioning. *Arthritis Care Res* 1995;8:94-101.
- 2 Kwok 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.
- 3 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.
- 4 Hidding A, van Santen M, De Klerk E, Gielen X, Boers M, Geenen R, et al. Comparison between self-report measures and clinical

- 5 observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. *J Rheumatol* 1994;21:818-23.
- 6 Jacobs JW, Oosterveld FG, Deurbouts 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 Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychol Bull* 1995;117:167-78.
- 8 Cooper H, Hedges L, eds. *The handbook of research synthesis*. New York: Russel Sage Foundation, 1994.
- 9 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.
- 10 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
- 11 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58.
- 12 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.
- 13 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.
- 14 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.
- 15 Carr AJ, Donovan JL. Why doctors and patients disagree. *Br J Rheumatol* 1998;37:1-4.
- 16 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.
- 17 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.

Accepted: 7 April 2008

## Patients' attitudes to the summary care record and HealthSpace: qualitative study

Trisha Greenhalgh,<sup>1</sup> Gary W Wood,<sup>1</sup> Tanja Bratan,<sup>1</sup> Katja Stramer,<sup>1</sup> Susan Hinder<sup>2</sup>

<sup>1</sup>Department of Primary Care and Population Sciences, University College London, Highgate Hill, London N19 5LW

<sup>2</sup>RaFT Consulting, Clitheroe, Lancashire BB7 4BN

Correspondence to: T Greenhalgh p.greenhalgh@pcps.ucl.ac.uk

BMJ 2008;336:1290-5  
doi:10.1136/bmj.a114

### ABSTRACT

**Objective** To document the views of patients and the public towards the summary care record (SCR, a centrally stored medical record drawn from the general practice record) and HealthSpace (a personal health organiser accessible through the internet from which people can view their SCR), with a particular focus on those with low health literacy, potentially stigmatising conditions, or difficulties accessing health care.

**Design** 103 semistructured individual interviews and seven focus groups.

**Setting** Three early adopter primary care trusts in England where the SCR and HealthSpace are being piloted. All were in areas of relative socioeconomic deprivation.

**Participants** Individual participants were recruited from general practice surgeries, walk-in centres, out of hours centres, and accident and emergency departments. Participants in focus groups were recruited through voluntary sector organisations; they comprised advocates

of vulnerable groups and advocates of people who speak limited English; people with HIV; users of mental health services; young adults; elderly people; and participants in a drug rehabilitation programme.

**Methods** Participants were asked if they had received information about the SCR and HealthSpace and about their views on shared electronic records in different circumstances.

**Results** Most people were not aware of the SCR or HealthSpace and did not recall receiving information about it. They saw both benefits and drawbacks to having an SCR and described a process of weighing the former against the latter when making their personal choice. Key factors influencing this choice included the nature of any illness (especially whether it was likely to lead to emergency care needs); past and present experience of healthcare and government surveillance; the person's level of engagement and health literacy; and their trust and confidence in the primary healthcare team and the