EBANDOLIER

The Bandolier electronic resource continues to grow and try and make itself more useful. You might lose your paper copies behind the sofa, but the electronic version can always be found. For the new year eBandolier has had a makeover and reorganisation. It also has plans to expand in 2001. The areas for expansion include, but are not restricted to:

Diagnostics and diagnosis, supported by the NeLH, sees additional abstracts of good evidence, and examples about how to think about tests and methods.

Migraine is now a major site, sponsored by the Gwen Bush Foundation and MSD, with many systematic reviews conducted especially for Bandolier, as well as information about how trials are conducted, their outcomes, diagnosis, and what patients think.

Healthy living grows using Bandolier’s own resources. In 2001 we would love to hear about practical methods people have found to influence us to change to healthier lifestyles.

Complementary and alternative therapy pulls together the best evidence we have found in this area, supported by the BUPA Foundation. More will come, plus help on how to think about issues like quality and validity.

New in 2001 are planned sites on atrial fibrillation, on vaccines, and on palliative care, because these are areas on which readers have asked us to concentrate.

Managing to make a difference. Supported by the NeLH and NHSE we plan a series of web essays by Michael Dunning (of GRIP, PACE and ImpAct fame) on how to make a difference. This month the web essay is on avoiding elephant traps. More stories from the literature an examples of good management in the NHS continue to be added. Below we tell you how you can contribute.

Sharing success

This month we have added several new pieces to the Bandolier/ImpAct website. Two case studies were sent to us because those involved wanted to share their work with colleagues and saw ImpAct as a good way of doing so. They are part of our growing post bag.

We do not have space to include all the material we receive in the paper version of ImpAct but we can post them on our website. It’s a good place to be: the hit rate is about 80,000 per week and rising. Our main criteria are: the change must be measurable, it should be affordable, and it must be transferable. If you want to join in, we will help you prepare the report for publication.

In this issue

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which migraine treatment strategy is most effective?</td>
<td>p. 1</td>
</tr>
<tr>
<td>Continuous quality improvement</td>
<td>p. 3</td>
</tr>
<tr>
<td>HRT and tooth loss</td>
<td>p. 4</td>
</tr>
<tr>
<td>Diagnosing acute sinus infections</td>
<td>p. 5</td>
</tr>
<tr>
<td>Herbal remedies for asthma</td>
<td>p. 6</td>
</tr>
<tr>
<td>Ranking quality of life in chronic disease</td>
<td>p. 7</td>
</tr>
<tr>
<td>Oxford Masters programme in EBM</td>
<td>p. 8</td>
</tr>
</tbody>
</table>

The views expressed in Bandolier are those of the authors, and are not necessarily those of the NHSE.

W HICH  MIGRAINE  TREATMENT STRATEGY IS MOST EFFECTIVE?

A number of strategies can be used to treat acute migraine attacks, each utilising some part of the evidence base.

For instance, the initial attack could be treated with aspirin or simple analgesic, and if or when that fails, a triptan could be used. That is a step strategy within an attack.

A different approach may be to try aspirin or simple analgesic for a few attacks. It will work for some, but for those for whom it does not work, a triptan may be an alternative treatment. That is a step strategy across attacks, and is probably the strategy most likely to be used in the UK as it is probably seen as the cheapest.

A third way would be to assess the individual patient for the severity of the disorder, and then to treat appropriately: mild disease might be treated with aspirin or simple analgesics, while more severe disease might be treated with a triptan. This would be stratified care.

It just so happens that a randomised controlled trial indicates that stratified care produces the best results [1].

Trial

The trial was randomised, but open-label, and examined multiple migraine attacks for patients with established diagnosis of migraine according to International Headache Society criteria. Patients completed the MIDAS questionnaire, one that measures lost time in three domains of activity [2]. Patients were assigned a grade of disability from I (little or infrequent disability), grade II (mild or infrequent disability), grade III (moderate disability) to grade IV (severe disability). Patients with grade II-IV disability were included.
Randomisation was to:

**Stratified care:** grade II patients received aspirin 800 to 1000 mg plus metoclopramide 10 mg for all six attacks. Those with grade III or IV received zolmitriptan 2.5 mg.

**Step care across attacks:** Patients treated the first three attacks with aspirin 800 to 1000 mg plus metoclopramide 10 mg. Those without adequate relief took zolmitriptan 2.5 mg for the next three attacks.

**Step care within attacks:** Patients treated all attacks with aspirin 800 to 1000 mg plus metoclopramide 10 mg first. If adequate relief was not obtained by two hours, they then took zolmitriptan 2.5 mg.

Patients had to have pain of moderate or severe intensity before taking their medicine. Two hours later, if their pain was gone, or of only mild intensity, this was considered a headache response. If the pain was gone, this was a "pain free" outcome. These are standard migraine outcomes, and other outcomes are available in the study.

### Results

In the three treatments groups, 1062 patients were randomised. Twenty percent of patients withdrew or were lost for various reasons, mostly innocuous. Only 3% withdrew because of an adverse event, and 0.2% because of deteriorating condition. Groups were well balanced.

More patients had a two-hour headache response in the stratified care strategy than for either step care strategy (Figure 1). About half the patients in the stratified strategy had a two hour headache response.

More patients were pain free at two hours in the stratified care strategy than for either step care strategy (Figure 2). About one in five patients in the stratified strategy was pain free at two hours.

Adverse events were equally common in all three groups, and were predominantly mild and transient. Adverse event study withdrawals were evenly distributed across the groups.

### Comment

Most guidelines probably use a step up approach, similar to that of step up across attacks, but with many more steps. Because of the time involved, and because of repeated failure of treatment, some patients simply become disenchanted and seek other forms of treatment. Treating the appropriate patient appropriately from the beginning is a better bet. It takes less time, is more effective, and is without the "hassle factor" for patient and doctor. This is exactly what evidence-based medicine was supposed to be about, and reading the definition of EBM in the context of this trial is rewarding. It is striking, though, that the best treatment strategy still means that four out of every five migraine sufferers who are treated still have some pain two hours after treatment started.

### References:

This is a thorny problem in healthcare. There’s not really an awful lot laid down in terms of what to do and how to do it. There’s not even much in the literature. A report of continuous quality improvement to reduce planned induction of labour from Vancouver [1] has interesting lessons, not only about how to do it, but also about how to think about it.

Background

The study was carried out at Canada’s busiest maternity hospital, with 7000 births a year and 125 nurses, 36 obstetricians and 124 family physicians. In the mid 1990s the induction rate was running at 23-25% of all births, and while there was little evidence about what was appropriate, this rate was thought to be too high. Initial attempts to draft guidelines and change a booking form foundered mainly due to a lack of clarity in guidelines and ownership of the systems: nurses were set to monitor doctors, and good working relations suffered.

Doing it properly

Doing it properly involved:

♦ A multidisciplinary team representative of all professional bodies and patients.
♦ Staff remunerated for extra time worked.
♦ An analysis of the problems and their definition.
♦ A critical review of the literature to obtain the best evidence, and the use of institutional consensus.
♦ Clarity about how the process was to be implemented using a new booking form.
♦ Peer review of all inductions by obstetricians and family physician, using a process that minimised hostility through consensus and use of best evidence, leading to eventual handing of the review process back to nurse team leaders.

Results

The number of inductions per 1000 deliveries fell after the new programme was introduced, for all indications (Table). The rate fell from 239 per 1000 deliveries (24%) to 183 per 1000 deliveries (18%). This is equivalent to the intervention preventing one induction for every 18 deliveries.

Comment

The Table deserves a lingering examination to see just where the fall in inductions occurred. The single largest category was one of logistics, essentially one of family convenience, and the second largest was of “other”, which included drug abuse with risk of elopement (sic). Over half the fall in inductions was those planned for non-medical reasons.

This provides a useful indication of how other units wanting to look at their induction rate might start. Go for the large numbers with simple answers. This group may be one such.

Overall, though, this sensitively written article demonstrates again the quality available in “how to do it” work. The importance here is that it compares a successful and unsuccessful approach.

References:

Table: Planned induction of labour per 1000 deliveries before and after instituting a continuous quality improvement programme

<table>
<thead>
<tr>
<th>Inductions per 1000 deliveries</th>
<th>15 months before implementation</th>
<th>9 months after implementation</th>
<th>Inductions saved per 1000 women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post term pregnancy</td>
<td>51</td>
<td>48</td>
<td>3</td>
</tr>
<tr>
<td>Premature rupture of membranes</td>
<td>53</td>
<td>51</td>
<td>2</td>
</tr>
<tr>
<td>Suspected foetal jeopardy</td>
<td>36</td>
<td>29</td>
<td>7</td>
</tr>
<tr>
<td>Pre-eclampsia</td>
<td>30</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>Maternal disease</td>
<td>20</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Family circumstance/place of residence</td>
<td>16</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Other</td>
<td>29</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Not recorded</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Total per 1000 deliveries</td>
<td>239</td>
<td>183</td>
<td>56</td>
</tr>
</tbody>
</table>
HRT AND TOOTH LOSS

Reasons for giving hormone replacement therapy to postmenopausal women are fairly well known. Around the menopause it is to do with control of symptoms, especially flushing. Later it is to more to do with protecting bone density to prevent osteoporosis and later fractures. One of the unintended consequences is the protection of teeth.

The arguments that protection of teeth is a real phenomenon are several, if predominantly indirect. Osteoporosis affects jaw bones as much as long bones and the spine. Loss of mineral from the area surrounding the teeth leads to tooth loss. Women with low bone mineral density in spine and hip bones have fewer teeth than those with higher bone density. A systematic review [1] seeks to pull together several links in the chain of evidence linking tooth loss and HRT, and to quantify the economic effects.

Review

The review used MEDLINE, other databases and manual searches for the period from 1980 to 1998. Eligible studies were full published papers of any study design of a minimum of three months looking at postmenopausal women and with a minimum of 10 patients per study arm. Only English language papers were used.

Subjects eligible were postmenopausal women receiving HRT or not and those with documented osteoporosis. Outcomes used were dental outcomes (predominantly use of dentures, number of teeth, or percentage edentulous), and HRT use.

Results

Dental outcomes

Twenty studies were eventually included, with 13,700 women with the eligibility and outcome criteria fulfilled. The main dental results are shown in the Table. For all postmenopausal women, with a mean age of 67 years, there was denture use in just under half (47%). In women who had not used HRT, this rose to 69%, but fell to 27% for those with a history of HRT use. Women with a history of HRT use had more teeth than those who had not used HRT.

The information on dentures and number of teeth in women with and without documented osteoporosis was limited to a few hundred women, and was not informative.

Economic analysis

Economic analysis used published costs of dental procedures (fillings, crowns, root canals, extractions, X-rays dentures) using US and UK costs to give a range of costs expressed in US$. The high and low costs were used to estimate a range of likely costs, and these were applied to a hypothetical cohort of 1000 women who were and were not HRT users, and these costs were applied to high and low estimates for full or partial dentures or tooth extractions. The period assumed HRT use for five years, which was the mean duration in the studies in the review.

The excess cost of dental treatment associated with 1000 non-HRT users ranged from $14,000 (£10,000) to $300,000 (£214,000), with a mean of about $100,000 (£71,000). This means that use of HRT saves about £71 a year on average in dental treatment for each woman treated.

Comment

Bandolier is old enough to remember when complete removal of teeth was seen as a useful 21st birthday present, or even as a wedding present for a blushing bride. How things change. Perhaps the effects of HRT on tooth retention are well known. The evidence from studies is limited, indirect (no randomised trials with this as the main outcome), and with different outcomes. Yet by pulling together evidence from different sources, and by doing some not very sophisticated analysis, the conclusion that HRT use prevents costs elsewhere begins to stand up.

It is a sort of unexpected consequence of HRT use. It benefits the system, and also women who otherwise might have to spend more time in the dentist’s chair, and be concerned about their appearance. Seventy years is not old for a woman, and retention of teeth may well bring even more unexpected additional benefits stemming from confidence and a refusal to be old.

References:


Table: Dental outcomes in women, and the effect of hormone replacement therapy (HRT)

<table>
<thead>
<tr>
<th>Women</th>
<th>Number of studies</th>
<th>Number of women</th>
<th>Mean age (years)</th>
<th>Percent with dentures (CI)</th>
<th>Number of remaining teeth (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All women</td>
<td>10</td>
<td>5757</td>
<td>67</td>
<td>47 (44)</td>
<td>20 (8)</td>
</tr>
<tr>
<td>HRT</td>
<td>4</td>
<td>2881</td>
<td>69</td>
<td>27 (28)</td>
<td>22 (7)</td>
</tr>
<tr>
<td>No HRT</td>
<td>8</td>
<td>2073</td>
<td>69</td>
<td>69 (43)</td>
<td>16 (7)</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>3</td>
<td>203</td>
<td>68</td>
<td>73 (68)</td>
<td>5 (10)</td>
</tr>
<tr>
<td>No osteoporosis</td>
<td>3</td>
<td>112</td>
<td>68</td>
<td>69 (78)</td>
<td>6 (10)</td>
</tr>
</tbody>
</table>
**Diagnostic sinuses is acute enough, usually diagnosed by clinical examination, and treated with a course of antibiotics. The trouble is that many of the symptoms of acute sinusitis overlap with other common nasal complaints, so misdiagnosis and over-treatment with antibiotics is thought to be common. Diagnostic interventions of more invasive or high technology nature are available, like X-ray, CT scanning, or sinus puncture. So when two systematic reviews of diagnostic tests for the same condition appear in the same journal, almost adjacent to one another [1,2], clarity may be expected. Unfortunately, it’s not quite like that, and another two brains problem rears its head. The clinical bottom line is that we don’t have adequate diagnostic methods.

**Reviews**

The review from Boston [1] used a MEDLINE search for studies that had to compare the ability of at least two tests to diagnose acute sinusitis. This review included some studies despite an inadequate description of eligibility criteria, for sensible reasons. The sensitivity and specificity of one test had to be compared with another, reference, test. The main analytical tool was the summary receiver operating characteristics curve, when sensitivity is plotted against 1-specificity.

The review from Helsinki, Oulu and Copenhagen [2] also searched MEDLINE, and a Finnish database, and hand searched four journals and contacted experts in the field. Studies with the reference standard of sinus puncture or CT scanning were accepted. Validity criteria from a Cochrane working group were used for validity, and excluded some studies that failed some of these criteria. Summary receiver operating characteristics were also used as an analytical tool, but likelihood ratios were also produced.

**Included studies**

Review 1 included 13 studies, six with patients from a hospital or emergency room setting, three from general practice, one from both, and three did not say how patients were recruited. Review 2 included nine studies, eight of which were included in review 1 (though there were differences between the reviews in, for instance, the year of publication of studies with otherwise identical references, as well as internal inconsistencies).

**Results**

Review 1 describes and criticises the studies and the methods, and gives overall curves for several different comparisons. Review 2 also criticises the included trials, and gives prevalence, sensitivity, specificity and likelihood ratios for each method compared with sinus puncture (Table). Using the weighted prevalence figure and the likelihood ratios, radiography, ultrasound and clinical examination fail to generate post-test probabilities of more than 80% or less than 20%.

**Comment**

These are cracking reviews, of high quality, and well written. Bandolier marginally prefers review 2 because it allows the calculation of post-test probabilities, and shows clearly how radiography, ultrasound and clinical examination fail to provide sufficiently high post-test probabilities to be truly useful: no better than 80% or 20% is probably not good enough. Review 1, though, has more details about the individual studies and the components of the diagnostic tests, so that is more useful in understanding what is going on.

Both have excellent discussions of the problems of diagnostic test reporting, and the particular problems about diagnosis of acute sinus infection, including those of sinus puncture, regarded as the gold standard. Taken together, these illuminate the problems of diagnostic test methods and quality, and their review. They both conclude, rightly, that it’s back to the drawing board. Review 2 additionally adds the visionary concept of combining the two aspects of clinical management, diagnosis and treatment, as a unit.

Diagnostic testing evidence is in short supply. The importance of this particular topic lies in the opportunities for research across the primary and secondary care boundaries, and combining clinical examination with high-tech approaches.

**References:**


<table>
<thead>
<tr>
<th>Comparison (with sinus puncture)</th>
<th>Number of studies</th>
<th>Number of patients</th>
<th>Weighted prevalence (% range)</th>
<th>Positive likelihood ratio</th>
<th>Negative likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiography</td>
<td>7</td>
<td>996</td>
<td>57 (26-80)</td>
<td>3.4</td>
<td>0.26</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>7</td>
<td>940</td>
<td>56 (27-80)</td>
<td>2.8</td>
<td>0.30</td>
</tr>
<tr>
<td>Clinical examination</td>
<td>2</td>
<td>245</td>
<td>46 (44-48)</td>
<td>3.3</td>
<td>0.40</td>
</tr>
</tbody>
</table>
HERBAL REMEDIES FOR ASTHMA

Overheard in a post office recently was a comment from one person to another, with the effect that a grandchild was being properly treated with natural herbal remedies for its asthma, and wasn’t the mother doing a good thing. Hard to bite one’s tongue, sometimes, but without the exact evidence to hand, or even with it, silence is golden.

Given that about four million people in the UK have asthma, the question about efficacy of alternative therapies will be a frequently-asked question. How can it be answered? A systematic review has found all the randomised trials of herbal remedies [1]. How should we think about evaluating that evidence to reach our own conclusions when the question is asked of us?

Review

Searching for studies was very thorough. It consisted of using a range of electronic databases, including the Cochrane Library, with a limit at end 1999. Other files and researchers were interrogated, No language limitation was imposed. Any randomised trial of any herbal remedy was included.

Asthma diagnosis allowed use of recognised criteria or reversible airway constriction. Outcomes were lung function tests, symptoms, use of medicines and events like healthcare use, days lost from work or school, or steroid use. In the review, lung function tests were used, with a change of 15% or more considered clinically relevant.

Results

Seventeen randomised trials were found, six using traditional Chinese medicines, eight using traditional Indian preparations, and three others. No two studies used the same herbal remedy in the same preparation, as best we can tell. Only nine of these trials were described as double blind, and the overall methodological quality was poor. Using a scale of up to five points for methodological quality [2], 10 trials scored two points or fewer, four scored 3 and three scored 4. Nine of the 17 trials reported a clinically relevant improvement in lung function scores.

Comment

This brief rehearsal of the results does not do justice to a fine review that summarises each of the seventeen studies and is thoughtful. The review concludes that there is no definitive evidence for any herbal preparation, and the discussion on the potential for harm is informed and informative. But there is an opportunity to use the results to hone our skills about how to think beyond the bare result, even if it is statistically significant. This means thinking about quality, bias, validity and relevance.

What we might want to do is to consider what criteria evidence from randomised trials has to meet before becoming acceptable. We know, for instance, that studies with a quality score of 2 or less tend to be more positive in their results, a form of quality bias. So studies may have to have quality scores of three or more. Then there is the validity question. Asthma is a chronic disorder, so for relevance we might want to use only studies that have a duration of at least four weeks.

If we use only those criteria, what do the results look like in terms of the number of studies that meet them? The Table shows this. The number of studies with quality score of at least 3, and a duration of four weeks or more is three, all traditional Indian remedies.

Only one of these has positive outcomes. This study, published in 1974, had 123 bronchial asthmatics (age unreported) enrolled and measured lung function at various times over 12 weeks. The peak benefits were at four weeks.

Now we could have chosen other criteria of validity, like size (small trials may be biased). Or we could have chosen year of publication (we want up-to-date information) or place conducted (concerned about geographical bias). But we don’t need them, and if we did use them, we’d get much the same answer.

The point, at the end, is this. The evidence for effectiveness for herbal remedies in asthma doesn’t amount to a hill of beans. Those who promote them do a disservice to consumers. We can be confident that no-one has a clue whether they work or not, and that no conventional medicine would be licensed with the evidence available.

References:

Table: Randomised trials of herbal remedies for asthma

<table>
<thead>
<tr>
<th>Type of herbal remedy</th>
<th>Total number of studies</th>
<th>Number with quality score 3 or more</th>
<th>And with duration 4 weeks or more</th>
<th>And have result favouring herbal remedy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Indian</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>1 *</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* note that this study, published in 1974, had peak benefit at four weeks
The question of which chronic disease most impacts upon quality of life can lighten many a dreary hour. Anecdote piles upon anecdote, but the problem is that the plural of anecdote is not data. What is needed is surveys using the same instrument to measure quality of life, used in large enough samples of patients, with a similar range of disease severity, and with a similar demographic base. Even then there may be problems in interpretation, but it might provide some better insights into disease impact on people. A study with about 15,000 patients from Holland gives us just this [1]. But prospective readers should beware. This is definitely a two-brains paper, but worth conquering.

**Study**

All research groups known to examine chronic diseases in the Netherlands were contacted to see what data sets were available. Studies had to use a standardised quality of life instrument, have full coverage of quality of life domains, include a range of chronic diseases, be big (at least 200 patients), have medically confirmed diagnoses, be obtained since 1992 and be geographically broad.

Eight data sets broadly fulfilling these categories were obtained, with information on about 15,000 people. They all used SF-36 or SF-24. These were analysed by quality of life dimension (physical functioning, physical role functioning, bodily pain, general health, vitality, social functioning and mental health) according to:

- Disease clusters (grouping together similar diseases. For instance musculoskeletal conditions of osteoarthritis, joint complaints, rheumatoid arthritis and back impairments).
- Disease categories (ranking the individual diseases within the cluster).
- Patient characteristics (sociodemographic variables, like age, gender, education).

The method used was the ranking of mean scores. Thus if three diseases scored (say) 5, 10, and 15 (with 5 the “best” score), then they would be ranked 1, 2 and 3. This was done for all quality of life domains, and the ranks for individual domains added together. This summed rank produces low scores for the diseases or disease clusters causing the least distress, and high scores for those causing the most problems.

**Results**

**Disease clusters**

The summed rank scores for chronic disease clusters are shown in the Figure. Musculoskeletal conditions, renal disease, cerebrovascular/neurological conditions and gastrointestinal conditions impacted most severely on quality of life.

---

**Figure: Summed rank scores for disease clusters. Higher scores imply poorer quality of life**

![Graph showing summed rank scores for various disease clusters](image-url)
Disease categories

Some examples are useful here. For instance, in musculoskeletal conditions, osteoarthritis had more adverse impact than back impairments, which scored higher (worse) than rheumatoid arthritis. For neurological conditions, Parkinson’s disease or epilepsy, multiple sclerosis and stroke scored higher than migraine or neuromuscular disease. For psychiatric disorders depression scored worse than anxiety which in turn was worse than alcohol problems.

Patient characteristics

Patients who were older, female, had a low level of education, were not living with a partner, and/or had at least one comorbid condition had the poorest quality of life.

Comment

There will be limits to how far these sort of data can be subdivided and still give us valid conclusions. So where there is the largest agglomeration of information is where the strongest conclusions lie. For this analysis, this is with the comparisons across disease categories.

Moreover, there are also issues within the quality of life measures that an overall ranking will not highlight. This will be in the difference between mental and physical functioning, for instance.

Many professionals will not be overly surprised at the ranking of disease clusters, or the categories within each cluster, or the conclusions regarding patient characteristics. But where there is limited information there will always be room for argument. Though this ranking exercise should not be over-interpreted, it does give us a firmer platform on which to base decision-making, and on which to base research efforts.

The man on the stair

But hang on a minute, advised the man on the stair. Do you really believe that musculoskeletal conditions cause a greater impact on quality of life than, say visual impairment? Does this result have face validity? Could it simply be wrong?

What might be the causes of a wrong answer? Well the combining of data in the meta-analysis may be incorrect, though the authors seem to have done a pretty fine job, and discuss in detail possible problems and why they are unlikely to occur. Then there are the original studies themselves used in the combining process. Issues of validity of the original study was part of the defining process of the analysis, which is why only eight of about 30 studies actually made it into the review. The others were excluded because they were judged inadequate.

Then there is the method used for measuring quality of life, in this case SF-36. This measurement tool has been around for ages, is much used, and for which there are manuals and methods written down in exquisite detail. Validity has been measured in lots of different ways, hasn’t it? And all that has been done so that we can be confident of the results obtained by using it.

The only other conclusion left would be that SF-36 is fundamentally flawed. That would have major repercussions, especially on all those discussions on things like cost per QALY that policy makers use for judging whether medicines can be bought by the NHS or other healthcare providers. Fundamental stuff, this, and why this paper [1] is so important about thinking about thinking about healthcare delivery.

References: