

# MELATONIN FOR THE PREVENTION AND TREATMENT OF JET LAG

Opher Caspi, MD, PhD

**Opher Caspi, MD PhD**, is a research assistant professor in the Program in Integrative Medicine at the University of Arizona. (*Altern Ther Health Med.* 2004;10(2):60-63.)

## CLINICAL SCENARIO

An otherwise healthy middle-age businessman who lives in California is planning an important business trip to Europe next month. Concerned about being too fatigued when he arrives there, he asks whether you have any tips on how he can avoid jet lag.

## SEARCHING FOR AN EVIDENCE-BASED ANSWER

Since medical decisions, whether diagnostic or therapeutic, are usually made within the specific context and particularities of the patient for whom evidence is being sought, the search for evidence must always start with the construction of a well-built clinical question that focuses on the specifics.<sup>1</sup> These questions are typically: who is the patient? What is the specific health issue to be addressed? What is the range of intervention options considered? And what is the set of potential outcomes desired or to be avoided? In the above case, the patient is an otherwise healthy middle-age man who would like to avoid jet lag while traveling eastward across many time zones. Please note how this case differs from, for example, a person who is already experiencing jet lag, or one that is about to travel westward a relatively short distance. Some other parameters, such as the patient's gender are likely to be inconsequential in this specific case, yet very important in other cases. Evidence-based medicine works best when the question includes just the important parameters. Too much specification could be a bit too confusing.

The next step is to retrieve the existing body of knowledge that pertains to the specific question under consideration. To maximize yield, it is important to prioritize the search so that you consult only those sources that are likely to provide you with valuable information. Sometimes textbooks would be sufficient (for example, when you want to learn about the natural history of a disease or its differential diagnosis). However, for the latest evidence on the efficacy and effectiveness of various therapeutics it is advisable to search for a systematic review that summarizes the evidence qualitatively, or better still, a meta-analysis that summarizes it quantitatively. A quick search of the Cochrane database reveals a Cochrane review "Melatonin for the prevention and treatment of jet lag,"<sup>2</sup> an abstract of which follows:

## ABSTRACT

**Background:** *Jet lag commonly affects air travelers who cross several time zones. It results from the body's internal rhythms being out of step with the day-night cycle at the destination. Melatonin is a pineal hormone that plays a central part in regulating bodily rhythms and has been used as a drug to re-align them with the outside world.*

**Objectives:** *To assess the effectiveness of oral melatonin taken in different dosage regimens for alleviating jet lag after air travel across several time zones.*

**Search Strategy:** *We searched the Cochrane Controlled Trials Register, MEDLINE, EMBASE, PsychLit and Science Citation Index electronically, and the journals 'Aviation, Space and Environmental Medicine' and 'Sleep' by hand. We searched citation lists of relevant studies for other relevant trials. We asked principal authors of relevant studies to tell us about unpublished trials. Reports of adverse events linked to melatonin use outside randomized trials were searched for systematically in 'Side Effects of Drugs' (SED) and SED Annuals, 'Reactions Weekly,' MEDLINE, and the adverse drug reactions databases of the WHO Uppsala Monitoring Centre (UMC) and the US Food & Drug Administration.*

**Selection Criteria:** *Randomized trials in airline passengers, airline staff or military personnel given oral melatonin, compared with placebo or other medication. Outcome measures should consist of subjective rating of jet lag or related components, such as subjective well being, daytime tiredness, onset and quality of sleep, psychological functioning, duration of return to normal, or indicators of circadian rhythms.*

**Data collection and analysis:** *Ten trials met the inclusion criteria. All compared melatonin with placebo; one in addition compared it with a hypnotic, zolpidem. Nine of the trials were of adequate quality to contribute to the assessment, one had a design fault and could not be used in the assessment. Reports of adverse events outside trials were found through MEDLINE, 'Reactions Weekly,' and in the WHO UMC database.*

**Main Results:** *Nine of the 10 trials found that melatonin, taken close to the target bedtime at the destination (10pm to midnight), decreased jet lag from flights crossing five or more time zones. Daily doses of melatonin between 0.5 and 5mg are similarly effective, except that people fall asleep faster and sleep better after 5mg than 0.5mg. Doses above 5mg appear to be no more effective. The relative ineffectiveness of 2mg slow-release melatonin suggests that a short-lived*

higher peak concentration of melatonin works better. Based on the review, the number needed to treat (NNT) is 2. The benefit is likely to be greater the more time zones are crossed, and less for westward flights. The timing of the melatonin dose is important: if it is taken at the wrong time, early in the day, it is liable to cause sleepiness and delay adaptation to local time. The incidence of other side effects is low. Case reports suggest that people with epilepsy, and patients taking warfarin may come to harm from melatonin.

**Reviewers Conclusions:** Melatonin is remarkably effective in preventing or reducing jet lag, and occasional short-term use appears to be safe. It should be recommended to adult travelers flying across five or more time zones, particularly in an easterly direction, and especially if they have experienced jet lag on previous journeys. Travelers crossing 2-4 time zones can also use it if need be. The pharmacology and toxicology of melatonin needs systematic study, and routine pharmaceutical quality control of melatonin products must be established. The effects of melatonin in people with epilepsy, and a possible interaction with warfarin, need investigation.

The abstract suggests that this Cochrane review may provide information that is pertinent to our clinical question. But how valuable is this information at the point of care? The practice of evidence-based medicine depends primarily on our ability to critically assess the validity and applicability of research evidence and successfully incorporate it into patient care. In what follows, I shall demonstrate how this can be done. In future inquiries, you can use the exact same set of questions to critically appraise other systematic reviews and meta-analyses.<sup>3</sup> The critical appraisal process of other forms of publication, such as randomized controlled trials, case studies, or studies that focus on diagnosis and prognosis rather than on therapeutics, requires a slightly different set of questions. A number of excellent sources are available to guide you in this process.<sup>4,6</sup>

#### Did the overview address a focused clinical question?

The authors of this Cochrane review did a fine job in defining their questions of interest. Specifically, they sought to evaluate 1) whether melatonin taken by mouth can prevent or alleviate jet lag associated with air travel across several time zones, 2) the evidence for the effectiveness of different dosage regimens, and 3) all suspected adverse effects of melatonin. Thus, the review has the potential to affect practice. Overall, this comprehensive set of questions serves as a good example for the value of systematic reviews and meta-analyses within the framework of the Cochrane collaboration.

#### Were the criteria used to select articles for inclusion appropriate?

Yes. Although only randomized trials were included in this review, this decision seems reasonable considering the relatively large number of potential confounders involved in determining the efficacy of a drug (melatonin) in a heterogenic setting such as air travel where the outcome (degree of jet lag) is subjective in nature. All key parameters were defined *a-priori*. Participants in these randomized trials were airline passengers, airline staff or

military personnel, so the results are potentially generalizable to large populations. The types of intervention examined were oral melatonin, compared with placebo or other medication, taken before, during and/or after travel. The primary outcome measure was subjective rating of jet lag, and components or correlates of this, such as fatigue, daytime tiredness, onset of sleep at destination, onset and quality of sleep, psychological functioning, duration of return to normal, and measures indicating the phase of circadian rhythms. These outcomes are relevant to the scope of our question.

#### Is it likely that important, relevant studies were missed?

This question is important because including only a portion of all available evidence may introduce systematic errors into the meta-analytic process and threaten its validity. Based on current standards for literature search<sup>7</sup> it appears unlikely that important relevant studies were missed in this case. The authors searched a large number of electronic databases using appropriate key terms, and supplemented it by hand searching of key relevant medical journals, and contact of principal authors so as to identify potential unpublished trials. One potential limitation, however, is that it remains unclear whether the search was limited only to research published in English. Although, there is no evidence that language restricted meta-analyses lead to biased estimates of intervention effectiveness, it is always a desirable practice to include all languages in the search strategy.<sup>8</sup> Also, since the original search covered all randomized clinical trials conducted and published only until 1999, it is desirable to check the literature for new evidence that may have emerged since the Cochrane review was conducted. The Cochrane Database of Systematic Reviews publishes updates on a regular basis. For assessing how current the evidence included in a Cochrane review is, look at the top of the review. In our case, the most recent substantive amendment to the review was made at the end of 2001 and a more minor amendment was made in late 2002.

#### Was the validity of the included studies appraised?

Since the extent to which systematic reviews and meta-analyses could guide healthcare decisions depends on the quality of the trials included (the famous “garbage in, garbage out” maxim), in all Cochrane reviews the quality of the original trials is assessed as a proxy measure of their validity. However, how to assess trial quality as part of a systematic review, or even if it should be assessed at all, remains uncertain (see sidebar). While proponents regard quality assessment as an important strategy to identify and reduce bias, opponents consider quality assessment as another source of potential bias.<sup>9</sup>

In the case of the melatonin Cochrane review, the validity of the included studies was only partially appraised. Although data related to methods, participants, interventions, and outcome were extracted from each eligible study, only 2 important parameters related to the studies’ validity were evaluated—allocation concealment and blinding. In clinical trials, potential biases fall into 4 categories that relate to systematic differences between comparison groups: (i) patient’s characteristics (selection bias), (ii) the provi-

sion of care apart from the treatment under evaluation (performance bias), (iii) the assessment of outcome (detection bias), and (iv) the occurrence and handling of patient attrition (attrition bias). Detailed accounts of these biases are provided elsewhere.<sup>10,11</sup>

Many of these potential sources of bias could not be thoroughly evaluated in the melatonin review since none of the original trial reports included a statement on allocation concealment, on how closely alike in appearance the test treatments were, what the participants were told about the trial they were entering, and what effects they would have been led to expect. All of these factors are important because prior expectations and subjects' perceptions could influence the effects and symptoms that they experience and report.<sup>12</sup> Also, none of the trial reports gave details of the source of the melatonin used and most did not state the pharmaceutical form used. These are all important components of trial validity, especially in the case of complementary and alternative medicine research.<sup>13</sup>

#### *Were assessments of studies reproducible?*

At every stage in performing a systematic review, reviewers must guard against potential bias or errors. This is most critical at the point at which studies are being selected for inclusion and when pertinent data is being extracted from original studies. Having multiple reviewers with slightly different perspectives assessing the data might help guard against bias and reduce error.<sup>7</sup> In our case, each author extracted data independently and differences were reconciled based on a rigorous methodology that was determined *a-priori*. The review includes tables that outline the characteristics of included studies.

#### *Were the results similar from study to study?*

Inferences about causal relationships between interventions (eg, melatonin) and outcomes (eg, prevention of jet lag) are prone to error. That is, researchers are always at risk of concluding that there is an effect, when, in fact, there is not (type I error), or concluding that there is no effect, when, in fact, there is (type II error). Multiple studies that examine the same question across different settings and populations provide multiple opportunities to test the causal hypothesis. Thus, the degree of consistency of findings across different studies serves as a crude indication of the extent to which we can trust the overall research findings. Studies that show inconsistent results should always prompt more investigation as to what accounts for the increased variance. On the other hand, studies that show consistent results increase the probability that the findings are to be trusted, given that there is no evidence of systematic error (bias). A general description of statistical methods used to examine between-trial differences can be found elsewhere.<sup>14</sup>

In the melatonin case, the results of the eight trials that were of adequate quality were remarkably consistent: they all showed that subjects who were given melatonin had better outcomes than subjects who were given a placebo. Two trials found no difference. These two trials assessed the different symptoms of jet lag as well as the rating of "jet lag" itself, and it might be

#### Assessing Methodological Quality

Although much controversy exists around what "quality of the data" exactly means, it is generally believed that the assessment of methodological quality provides a crude indication of the likelihood that the results of a clinical trial are a valid estimate of the truth.<sup>20</sup> The pros and cons of using one strategy or another when deciding whether or not to include individual trials of various methodological qualities in the processes of meta-analyses or systematic reviews remains largely a methodological conundrum. This uncertainty is because the aggregation of heterogeneous trials may not only be clinically inappropriate if the research question is not well specified, but also because it may introduce bias and/or variability to the overall analysis.<sup>21</sup> Analysis that is restricted to only high-quality trials, on the other hand, should result in better and more realistic estimates of treatment effects and hopefully greater acceptance of these results within the healthcare community.<sup>22</sup> But is this really the case? Do we have the appropriate methods to separate the wheat from the chaff?

On face value one of the most important dimensions of methodological quality is validity. Whereas some scholars advocate limiting the scope of methodological quality to only internal validity,<sup>23</sup> others consider both internal and external validity, in conjunction with statistical analysis as a more appropriate framework for assessing methodological quality.<sup>11</sup> Limiting methodological quality to internal validity only, defined here as "the confidence that the trial design, conduct, analysis, and presentation has minimized or avoided biases in its intervention comparisons" not only results in the exclusion of other methodological aspects of quality, such as precision and reliability of measures, but it also does not address other pertinent issues, such as ethics and scientific merit.<sup>20, 24</sup> The more inclusive approach to methodological quality, on the other hand, which refers to the construct as "the overall validity of the findings of each trial, taking into account all aspects of design and statistical interpretation which have bearing on the accuracy of the efficacy estimate,"<sup>25</sup> is likewise not free of some vagueness and methodological concerns.

To illustrate how subjective quality assessment can be, consider the discrepancy between the Cochrane review and another reliable source of evidence-based medicine<sup>26</sup> with relation to one of the trials included in the melatonin systematic review. Thus, whereas trial validity (both internal and external) seems to be at the heart of the construct of methodological quality, it should be remembered that validity is often distinct from quality. As Smith et al<sup>25</sup> correctly point out, quality can be assessed using a scale or a checklist; however, that does not necessarily assure at all that a trial is of adequate design to answer the question it poses. Consider the following example. A trial with a high quality score (ie, properly randomized and double-blind) would not be valid if the trial investigated patients in a manner that did not mirror reality. Another major problem arises from the assessment process itself, which requires judgment of the quality of individual clinical trials. In most instances the only way to assess the quality of a trial is to rely exclusively on the information contained in its report. It is thus important to distinguish further between assessing the *quality of a trial* and the *quality of its report*.

that not all symptoms change at the same rate and that the time of assessment might be important. In other words, jet lag might be interpreted differently at different times of the day.<sup>15</sup>

**Were all clinically important interventions and outcomes considered?**

Unfortunately, no trials have directly assessed the use of melatonin in conjunction with other strategies, such as light control, and only one trial compared melatonin to a hypnotic. Thus, it remains unknown whether an additive or synergistic effect occurs when a more comprehensive approach to jet lag prevention and treatment is taken. Likewise, it remains unknown whether melatonin is useful and safe in children and in old people, and if so how it should best be used. Close examination of the data extracted from each trial led the authors to conclude that for many of the outcomes of interest (eg, sleep latency and quality) results cannot be combined because the methods of measurement and reporting differed.

**What are the overall results of the review?**

Melatonin, taken close to the target bedtime at the destination (10pm to midnight), decreases jet lag from flights crossing five or more time zones. Daily doses of melatonin between 0.5 and 5mg are similarly effective, except that people fall asleep faster and sleep better after 5mg than 0.5mg. Doses above 5mg appear to be no more effective. The relative ineffectiveness of 2mg slow-release melatonin suggests that a short-lived higher peak concentration of melatonin works better. The benefit is likely to be greater the more time zones are crossed, and less for westward flights. The timing of the melatonin dose is important. If it is taken at the wrong time, early in the day, it is liable to cause sleepiness and delay adaptation to local time. The incidence of other side effects is low. Case reports suggest that people with epilepsy, and patients taking warfarin may have adverse effects from melatonin.

**How precise were the results?**

Since medical research is almost always done on samples, rather than on populations, the true effect of a treatment in the entire population is often unknown. What we have instead, is an efficacy estimate provided by rigorous controlled trials. This estimate is called a point estimate. The point estimate reminds us that, although the true value lies somewhere in its neighborhood, it is unlikely to be precisely correct.<sup>4</sup> To assess what this “neighborhood” is like, we use a statistical strategy called confidence intervals that represent a range of values within which we can be confident that the population parameter lies. That is where precision comes in. The narrower the confidence intervals are, the more confident we can be that the point estimate is precise with respect to the population parameter, and vice versa. Therefore, it is always a good practice to look not just at statistical significance tests (*P* values), but also at the confidence intervals that surround the point estimate. An excellent account of confidence intervals and their utility is provided elsewhere.<sup>16</sup>

In the melatonin Cochrane review, the authors attempted to provide a point estimate that represents the potential efficacy of

**Understanding the Numbers: Numbers Needed to Treat (NNT)**

The concept of NNT was reported first by Laupacis et al,<sup>27</sup> in 1988 with the intent to provide readers with additional information to help them decide whether a treatment should be used. The NNT indicates how many patients on average have to be treated to avoid an undesirable event or to achieve one desirable event at a specific point in time. It is calculated as the inverse of the absolute risk reduction (ARR) caused by treatment (Figure 1). Absolute risk reduction and NNT vary with the patient’s baseline risk of the target event. For patients at very high risk of the target event, the NNT will tend to be low, and treatment is likely to be justified. For patients at very low risk of the target event, NNT is likely to be high enough to raise doubts about whether treatment is warranted, even when the outcome being prevented is serious.

The 2 x 2 Table		Outcome	
		Yes	No
Exposure	Yes	a	b
	No	c	d

  

Relative Risk (RR)	=	$\frac{a/(a+b)}{c/(c+d)}$
Relative Risk Reduction (RRR)	=	$\frac{c/(c+d) - a/(a+b)}{c/(c+d)}$
Absolute Risk Reduction (ARR)	=	$\frac{c}{c+d} - \frac{a}{a+b}$
Number Needed to Treat (NNT)	=	$\frac{1}{ARR}$
Odds Ratio (OR)	=	$\frac{a/b}{c/d} = \frac{ad}{cb}$

Despite the intuitive appeal of NNT, concerns have been expressed about its limitations. Cook and Sackett,<sup>28</sup> noted that NNT presents a problem when the results of an RCT with patients at one baseline risk are applied to a particular patient at a different risk. Chatellier et al<sup>29</sup> expressed concern on extrapolating NNT to time points not considered in trials. Artalejo et al<sup>30</sup> pointed out that although the NNT for each type of patient can be narrowed down somewhat depending on his or her clinical characteristics, even then it is not possible to ascertain which patients will benefit, just as it is impossible to ascertain which patients will have adverse effects of exposure to risk factors for many chronic diseases because of individual variability. They expressed a concern that the use of NNT in clinical practice could lead to a decrease in therapeutic acceptance and compliance, if the patient’s perception of the number is different from that of the clinician.

Wu and Kottke<sup>31</sup> provided an excellent example to illustrate the problem with interpreting NNT. They reported a comparison of the NNT values for 3 interventions: cardiac transplantation, implantable cardioverter defibrillators (ICDs), and cholesterol lowering. The 1-year NNT was 1 for cardiac transplantation, 4 for ICDs, and at least 600 for a treatment that lowered serum cholesterol levels by 10%. On the basis of NNT, one might conclude that cardiac transplantation is the most effective way to control cardiovascular disease. However, they also estimated the potential contribution of each of the 3 interventions to a population-wide reduction in cardiovascular mortality. According to the demonstrated efficacy of each intervention and the candidacy rates in a population, cardiac transplantation could reduce cardiovascular mortality by 0.9%, ICDs could reduce cardiovascular mortality by 1.1%, and cholesterol-level decreases of 10% could reduce cardiovascular mortality by at least 4.8% and perhaps by as much as 7.8%.

melatonin compared to placebo in prevention and treatment of jet lag. To do so, they used a parameter that is often used in Cochrane reviews called weighted mean difference (WMD), which expresses the intervention effects as the weighted difference in mean values between the intervention and control groups. Using a jet lag severity scale of 0-100 where 0 means no jet lag and 100 means most severe jet lag, they found that the weighted mean score after melatonin was 25, and after placebo 48. When they used meta-analytic techniques, which allow for the differing variances in the trials, the difference between the two groups was even bigger (WMD 37.3; 95% CI 39.8-34.9). This is highly significant, both statistically and practically.

They then looked at the two trials that reported results for individuals (and not merely group means), and found that 16 out of 24 people (67%) given placebo experienced jet-lag after an eastward transatlantic flight, while only 4 out of 22 (18%) did so after 5mg melatonin. On this basis, they concluded that one of every two people taking melatonin would benefit. When results are presented in this way, we call it number needed to treat or simply NNT. The smaller the number, the more robust the effect of the intervention is. Despite the intuitive sense that clinicians have for NNT, an extreme caution needs to be practiced when NNT is used to convey effect size (see sidebar.) Nonetheless, an NNT of 2 generally means a huge effect.

#### **Are the benefits worth the harms and costs?**

Since the pharmacology and toxicology of melatonin have not yet been systematically studied it is not possible at this time to cast firm conclusions about its safety. Systematic review of the literature suggests potential interaction between melatonin and vitamin K antagonists such as warfarin, and a possible relationship between melatonin and seizures. Another potential source of harm has to do with the degree of purity of commercial preparation of melatonin. Independent analyses of some of these products purchased in health food stores in the US found some impurities, including lead.<sup>17, 18</sup> But, how to deal with this direct result of the 1994 Dietary Supplement Health and Education Act is the subject of an intense debate.<sup>19</sup>

#### **Can the results be applied to my patient care?**

Yes. The results of this systematic review indicate that when traveling across a number of time zones melatonin is an effective treatment for prevention and treatment of jet lag when used at bedtime in the day of travel and for up to four days after arrival. The authors, however, are right to point out that individuals differ greatly in the experience of jet lag, with some travelers extremely affected, while others report no jet lag symptoms. This suggests that individual differences may strongly influence the effectiveness of melatonin. It is concluded, therefore, that this information along with some other non-drug sleep hygiene tips<sup>15</sup> should be shared with the patient. Bon voyage!

#### **References**

1. Klassen TP, Jadad AR, Moher D. Guides for reading and interpreting systematic reviews: 1. Getting started. *Archives of Pediatrics & Adolescent Medicine* 1998;152:700-04.
2. Herxheimer A, Petrie KJ. Melatonin for the prevention and treatment of jet lag (Cochrane Review). In: The Cochrane Library, Issue 3, 2003. Oxford: Update Software.
3. Oxman AD, Cook DJ, Guyatt GH, for the Evidence-Based Medicine Working Group. User's guide to the medical literature, VI: how to use an overview. *JAMA* 1994;272:1367-71.
4. Guyatt G, & Rennie, D., editor. *User's guides to the medical literature*. Chicago: AMA Press; 2002.
5. Gehlbach SH. *Interpreting the medical literature*. 3rd ed. New York: McGraw-Hill, Inc.; 1993.
6. Geyman JP, Deyo RA, Ramsey SD, editors. *Evidence-based clinical practice: concepts and approaches*. Boston: Butterworth-Heinemann; 2000.
7. Jadad AR, Moher D, Klassen TP. Guides for reading and interpreting systematic reviews: II. How did the authors find the studies and assess their quality? *Archives of Pediatrics & Adolescent Medicine* 1998;152:812-17.
8. Moher D, Pham B, Klassen TP, et al. What contributions do languages other than English make on the results of meta-analysis? *Journal of Clinical Epidemiology* 2000;53:964-72.
9. Caspi O. Evidence-based complementary and alternative medicine: Critical appraisal of quality as a variable in meta-analyses and systematic reviews [Master's Thesis]: University of Arizona; 2001.
10. Jüni P, Altman, DG, Egger M. Assessing the quality of controlled clinical trials. *British Medical Journal* 2001;323:42-6.
11. Jüni P, Altman DG, Egger M. Assessing the quality of randomized controlled-trials. In: Egger M. SG, Altman DG., editor. *Systematic Reviews in Health Care*. London: BMJ Publishing Group; 2001. p. 87-108.
12. Shapiro AK, Shapiro E. How blind is blind? In: Shapiro AK, Shapiro E, editors. *The powerful placebo: From ancient priest to modern physician*. Baltimore: The Johns Hopkins University Press; 1997. p. 190-216.
13. Caspi O, Rowland-Seymour A. Gingko: smart pill or not? *Journal of Family Practice* 2003;52:393-94.
14. Moher D, Jadad AR, Klassen TP. Guides for reading and interpreting systematic reviews: III. How did the authors synthesize the data and make their conclusions? *Archives of Pediatrics & Adolescent Medicine* 1998;152:915-20.
15. Herxheimer A, Waterhouse J. The prevention and treatment of jet lag. *BMJ* 2003;326:296-97.
16. Braitman LE. Confidence intervals assess both clinical significance and statistical significance. *Annals Int Med* 1991;114:515-17.
17. Consumerlab.com. Product review: Melatonin supplements. Available at: <http://www.consumerlab.com/results/melatonin.asp> (Accessed September 15, 2003).
18. Anonymous. Melatonin. *Medical Letter* 1995;37:111-12.
19. DeAngelis CD, Fontanarosa PB. Drugs alias dietary supplements. *JAMA* 2003;290:1519-20.
20. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. *International Journal of Technology Assessment in Health Care* 1996;12(2):195-208.
21. Detsky AS, Naylor CD, O'Rourke K, McGeer, AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *Journal of Clinical Epidemiology*. 1992;45(3):255-65.
22. Moher D, Jadad, AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials* 1995;16:62-73.
23. Moher D, Pham, B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352(9128):609-13.
24. Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi List: A Criteria List for Quality Assessment of Randomized Clinical Trials for Conducting Systematic Reviews Developed by Delphi Consensus. *Journal of Clinical Epidemiology* 1998;51(12):1235-41.
25. Smith LA, Oldman AD, McQuay HJ, Moore RA. Teasing apart quality and validity in systematic reviews: an example from acupuncture trials in chronic neck and back pain. *Pain* 2000;86:119-32.
26. Bandolier. Melatonin for jet lag? Available at: <http://www.jr2.ox.ac.uk/bandolier/band82/b82-4.html> (Accessed: September 15, 2003).
27. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine* 1988;318:1728-33.
28. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *British Medical Journal* 1995;310:452-4.
29. Chatellier G, Zapletal E, Lemaitre D, et al. The number needed to treat: a clinically useful nomogram in its proper context. *British Medical Journal* 1996;312:426-9.
30. Artalejo FR, Banegas JR, Artalejo AR, Guallar-Castillón P. Number-needed-to-treat to prevent one death. *Lancet* 1998;351:1365.
31. Wu LA, Kotkote TE. Number needed to treat: caveat emptor. *Journal of Clinical Epidemiology* 2001;54(111-6).