adult patients, there would be approximately 800 to 1,000 patients in one practice to be screened to find approximately 200 subjects likely to have COPD. It would take 8 to 10 years to make a diagnosis in all of them. The question remains how to persuade a busy family doctor to perform such an additional work for a very long period of time. A study of patients in a university ambulatory health-care system found underutilization of spirometry for patients with respiratory symptoms compatible with COPD. Only 42% of such patients underwent spirometry. In the National Health and Nutrition Examination Survey III study, >80% of subjects with respiratory symptoms had visited a physician during the previous 12 months but did not undergo spirometry.

High-risk population screening programs are targeting smokers who do not see their family physician for years. The costs of such programs are also low. In our program, spirometry combined with an antismoking advice cost $8 (US dollars) per person screened.

Early diagnosis seems to be an easier part of the programs aiming to reduce morbidity and mortality from COPD. Much more difficult is to make a smoker with early COPD to stop smoking. The current treatments are still unsatisfactory (my view, to give the same significance and to pool together the “treatment failures” during the first 3 days and to pool together the relapses at 7 days and after is a mistake. While the first pooling might be due to specific triggers, the latter could have a more complex causality, including noncompliance.

In the face of significant heterogeneity, reviewers need to be mindful; they should refrain from the statistical pooling of data and should refine their search, focusing more on the exploration of patient characteristics and diagnostic evolution. Chest 2000; 118:981–989

References


A Difficult Step in Meta-analysis

Refining the Search

To the Editor:

I found the meta-analysis performed by Edmonds et al (June 2002) to be very instructive, but in some areas it needs to be refined.

The inclusion in the meta-analysis of patients who had been treated with a combination of corticosteroids (CS) and long-acting β2-agonists clearly should have been avoided in a meta-analysis in which the primary outcome was the relapse rate. It has been demonstrated that the rate is lower when using a combination of long-acting β2-agonists and CS, compared with studies using short-acting β2-agonists and CS.

Some of the outcomes described are too heterogeneous. In my view, to give the same significance and to pool together the “treatment failures” during the first 3 days and to pool together the relapses at 7 days and after is a mistake. While the first pooling might be due to specific triggers, the latter could have a more complex causality, including noncompliance.

In the face of significant heterogeneity, reviewers need to be mindful; they should refrain from the statistical pooling of data and should refine their search, focusing more on the exploration and description of the sources of heterogeneity.

Mihai-Sergiu Jalba, MD, PhD
Brooklyn Hospital
Brooklyn, NY

References


To the Editor:

We are pleased to respond to the points raised by Dr. Jalba about our article in CHEST (June 2002). First, we feel the concern about long-acting β-agonists (LABAs) is misleading for

Communications to the Editor
a number of reasons. It is unlikely that any of the trials included in the review would employ therapy with the combination of LABAs and corticosteroids (CS). However, even if they had, one would imagine the distribution would be similar across the groups in a randomized controlled trial, and their effect would be of limited concern. The trials in the review are older, and these drugs are infrequently used, even today. Furthermore, Dr. Jalba’s claim of benefit from LABAs only relates to the treatment of patients with chronic asthma, since there is no evidence that adding LABAs to therapy for patients with acute asthma is beneficial.

When trials using similar designs (ie, randomized controlled trials), comparisons (ie, inhaled CS vs CS), populations (ie, acute asthma), and outcomes (ie, relapse) exist, pooling is justified. In our review, there was no significant clinical or statistical heterogeneity, pooling was clearly appropriate. If the results of pooling demonstrate heterogeneity, then limited exploration is warranted, and pooling may not be appropriate. Conversely, searching for subgroup differences has been shown to generate erroneous results and should be considered far more carefully than Dr. Jalba would suggest. We further believe that the timing of relapse needs to be more fully evaluated before assigning the blame to patients for “late relapses.” We clearly appreciate the fact that the timing of relapse is important. That is precisely why we separated the outcomes based on time in the review. Finally, we challenge Dr. Jalba on the need for a better search and more exploration of heterogeneity. Cochrane reviews pride themselves on comprehensive, exhaustive (even exhausting!) searches to identify all available evidence on a focused question. Precisely how would Dr. Jalba suggest that the current search (which included searching for published, unpublished, English, and foreign language literature using a variety of well-accepted resources) could be more comprehensive? The recommendation to focus more on the sources of heterogeneity is a dangerous one, and we would strongly warn against it unless there is the strict application of accepted guidelines. Such fishing expeditions are to be strongly discouraged.

Marcia L. Edmonds, MD, MSc
Brian H. Rowe, MD, MSc
Division of Emergency Medicine, University of Alberta
Edmonton, AB, Canada

This research has been supported by the Division of Emergency Medicine, University of Alberta. Dr. Rowe was supported by a salary award from the Canada Research Council as the Chair of Emergency Airway Diseases (Ottawa, ON).

REFERENCES

**Sleep Apnea Devices and Sleep Apnea Surgery Should Be Compared on Effectiveness, Not Efficacy**

To the Editor:

Walker-Engström et al stated in their article in *CHEST* (March 2002) that treatment with a dental appliance “showed a significantly higher success and normalization rate” and thus had “superior effectiveness” than uvulopalatopharyngoplasty (UPPP) for the treatment of obstructive sleep apnea. Although this study had many strengths, such as randomization, long follow-up, and blinded sleep study interpretation, the conclusion depended on a key methodological fallacy that is common in studies such as this. I offer a quantitative analysis to show that their conclusion is not just “partly invalidated” by suboptimal compliance but, rather, is completely invalidated.

It is critical to distinguish *efficacy* (ie, the effect in the laboratory under ideal conditions) from *effectiveness* (ie, the effect in everyday life) when comparing surgical and nonsurgical treatments for sleep apnea. Effectiveness is far more relevant than efficacy to clinical practice. A perfect laboratory value with the device on (ie, efficacy measure) is of no help to the patient who does not wear the device outside the laboratory. Treatment adherence is not an issue with surgical therapy. Thus, it makes sense to compare only the effectiveness of surgical and nonsurgical treatments. It is a fallacy to compare the efficacy. Rather than exclude subjects who failed to receive or use their randomly assigned treatment, the authors should have counted those subjects as treatment failures in an assessment of effectiveness, as is required in a true intention-to-treat analysis.

Including these treatment failures reduces the 4-year apnea index success rate of the dental appliance from 81% (32 patients) to 54% (48 patients), and that of the UPPP from 53% (40 patients) to 49% (43 patients). The difference between treatment groups is no longer statistically significant (p = 0.68 [Fisher exact test]). Likewise, both the 4-year apnea-hypopnea index success rate and the normalization rate are no longer statistically different (p = 0.20 and p = 0.28, respectively). The results are similar even if we exclude the subjects who dropped out due to unrelated medical problems (p = 0.67, p = 0.13, and p = 0.27, respectively). Furthermore, the sleep study values obtained while patients wore the dental appliance in the laboratory should be corrected for the actual usage in everyday life in order to measure treatment effectiveness.

The fact that patient satisfaction (which is an inherent measure of effectiveness rather than a measure of efficacy) with UPPP was as high as with the dental appliance (satisfied subjects: UPPP, 30% [18/60]; dental appliance, 30% [18/60]) is irrelevant than efficacy to clinical practice. A perfect laboratory value with the device on (ie, efficacy measure) is of no help to the patient who does not wear the device outside the laboratory. Treatment adherence is not an issue with surgical therapy. Thus, it makes sense to compare only the effectiveness of surgical and nonsurgical treatments. It is a fallacy to compare the efficacy. Rather than exclude subjects who failed to receive or use their randomly assigned treatment, the authors should have counted those subjects as treatment failures in an assessment of effectiveness, as is required in a true intention-to-treat analysis.

Including these treatment failures reduces the 4-year apnea index success rate of the dental appliance from 81% (32 patients) to 54% (48 patients), and that of the UPPP from 53% (40 patients) to 49% (43 patients). The difference between treatment groups is no longer statistically significant (p = 0.68 [Fisher exact test]). Likewise, both the 4-year apnea-hypopnea index success rate and the normalization rate are no longer statistically different (p = 0.20 and p = 0.28, respectively). The results are similar even if we exclude the subjects who dropped out due to unrelated medical problems (p = 0.67, p = 0.13, and p = 0.27, respectively). Furthermore, the sleep study values obtained while patients wore the dental appliance in the laboratory should be corrected for the actual usage in everyday life in order to measure treatment effectiveness.

The fact that patient satisfaction (which is an inherent measure of effectiveness rather than a measure of efficacy) with UPPP was as high as with the dental appliance (satisfied subjects: UPPP, 30% of 46 subjects; dental appliance, 27 of 46 subjects) suggests that these treatments have similar effectiveness. In fact, the authors have reported previously that quality-of-life contentment (another inherent measure of effectiveness) improved significantly more for patients in the UPPP group than for those in the appliance group.

The suboptimal effectiveness of each individual treatment for...