Evidence-Based Amblyopia Treatment: Results of PEDIG Studies

David K. Wallace, M.D., M.P.H.

ABSTRACT

Retrospective studies without control groups are sometimes useful for suggesting treatment effects and generating hypotheses, but they suffer from bias and confounding. In response to the need for better evidence for treatment, the Pediatric Eye Disease Investigator Group (PEDIG) has conducted several randomized amblyopia treatment trials over the past several years. Results have shown that spectacles alone are a powerful treatment for amblyopia; patching is superior to spectacles alone; initiating fewer hours of prescribed patching seems to be as effective as traditional treatment; patching is effective in older children, particularly if they have not previously been treated; atropine is as effective as patching after six months; and weekend atropine is equally effective as daily atropine. All studies have limitations, but randomized clinical trials limit bias and confounding and thus provide the highest level of evidence of treatment effect.

AMBLYOPIA STUDIES IN THE PAST

Until recently, most published studies on amblyopia treatment were retrospective reviews of large numbers of patients. Many of these studies reported excellent results with patching for amblyopia. However, there are problems with applying data from such studies to clinical practice. For example, many patients in these retrospective studies were lost to follow-up and therefore excluded from the data analysis. This creates a selection bias because failure to follow-up can be strongly associated
with lack of compliance and lack of improvement with treatment. Therefore, when these patients are excluded, the remaining patients are more compliant with treatment and have better outcomes than the general population. A second problem with retrospective studies has occurred when visual acuity is measured using several different techniques. It is difficult to know if a child who was started on amblyopia treatment based on a fixation preference has truly improved when the outcome measure is done by Allen pictures. A third problem is that definitions of success have varied widely. In some studies, eligibility criteria and the definition of success have overlapped, such that patients could actually fail to improve with treatment but still be considered a treatment success. Finally, the effects of other treatments such as spectacles have not generally been taken into account in these retrospective studies.

**PAUCITY OF CLINICAL TRIALS**

Until recently, there were few clinical trials on amblyopia treatment; therefore, practice patterns were based on opinions of experts. This resulted in great variability in treatment guidelines due to the existence of several schools of thought. Ophthalmologists tended to base their amblyopia treatment recommendations on practice patterns of experts whom they encountered during training.

Why have there been so few clinical trials in amblyopia until the past few years? The relative expense of a clinical trial requires that it addresses a “burning” question. In addition, investigators must have equipoise, meaning that they have genuine uncertainty as to which of the treatment arms is superior. Clinical trials cannot typically be done to investigate interventions for rare conditions, because a large number of patients are generally necessary in order to have adequate power to show differences between treatment groups. Disadvantages of clinical trials include their relative expense, the time that they take to complete, and challenges of patient recruitment and retention.

**ADVANTAGES OF RANDOMIZED CLINICAL TRIALS**

The advantages of a properly designed and executed clinical trial are great. Randomization is a powerful technique to limit the potential for confounding. Confounding is a major problem with many nonrandomized studies, and it occurs when an observed association between two variables is due at least in part to a third variable. An example is the association between alcohol and oral cancer. If one were to determine the incidence of oral cancer in patients who drink versus those who do not drink alcohol, the initial conclusion might be that alcohol is strongly associated with oral cancer. However, in this case, the confounder that has been ignored is smoking. A confounder is associated with the exposure, associated with the outcome, and is not on the causative pathway between exposure and outcome. Smoking is associated with alcohol use, strongly associated with oral cancer, and is not on the causative pathway between the two (i.e., alcohol use does not cause smoking). In this case, if smokers and nonsmokers were analyzed separately, or if the analysis adjusted for effect of smoking, then the association between alcohol and oral cancer would appear to be much weaker. Randomization controls confounding by distributing both known and unknown confounders in roughly equal proportions between the treatment groups.

In addition to controlling confounding, a randomized clinical trial has other advantages such as limiting the potential bias of treatment assignment. For example, an investigator might want to compare outcomes in patients treated with atropine versus those treated with patching within
his/her practice. However, if patients are not randomized, then there are likely to be biases present that have resulted in some patients being treated with patching and others receiving atropine. Clinical trials also allow standardization of the intervention, prospective data collection, mask outcome measures, and investigation of adverse events and the effect of treatment on secondary outcomes. Data from control groups are useful to investigate the placebo effect and to study the natural history of a disease. Thomas Chalmers, who is considered by many to be the father of clinical trials, wrote that a carefully designed therapeutic trial is not only more scientific but also more ethical than the treatment of consecutive patients with a new drug of unproved value. He added that one has only to review the graveyard of discarded therapies to realize how many patients would have benefited by being assigned to a control group.

AMBYLOPIA QUESTIONS ADDRESSED BY PEDIG

Much of the information we have gained in the past few years has been from studies conducted by the Pediatric Eye Disease Investigator Group (PEDIG), a network of community and university based providers operating in over one hundred sites. PEDIG studies are funded by the National Eye Institute, and the data coordinating center is located at the JAEB Center for Health Research in Tampa, Florida. PEDIG studies use a standard visual acuity testing protocol. For children less than age 7, HOTV letters are presented individually with surround bars to account for the crowding phenomenon in amblyopia, and older children use the electronic ETDRS vision test.

Randomized clinical trials have been performed by the PEDIG to address several questions in amblyopia treatment:

1. How well do glasses alone treat amblyopia?
2. Do we really know that patching works?
3. How many daily hours of prescribed patching are necessary?
4. What happens when patching is stopped?
5. Does patching work in older children?
6. Does atropine work as well as patching?
7. How often does atropine need to be used?

HOW WELL DO GLASSES ALONE TREAT AMBLYOPIA?

The spectacle phase of Amblyopia Treatment Study 5 (ATS5) sought to determine the incidence of successful treatment of previously untreated anisometropic amblyopia with spectacles alone. Eligible children were 3 to nearly 7 years old, able to cooperate for the ATS visual acuity testing protocol, and had visual acuity at baseline between 20/40 and 20/400 with at least three lines of interocular difference. Amblyopia was secondary to strabismus, anisometropia, or both. Ten to 30 minutes after wearing spectacles for the first time, study participants had their baseline visual acuity tested. Follow-up visits then occurred every five weeks as long as amblyopia continued to improve or until it resolved, defined as visual acuity equal to or better than the acuity of the sound eye.

The study enrolled from 34 sites 84 patients who had previously untreated anisometropic amblyopia. Their mean age was 5.2 years and their mean amblyopic eye visual acuity was 20/80. After five weeks, patients had improved an average of almost two lines, and 59% had improved two lines or more. Many patients continued to improve for several weeks. Eighty-one patients completed the 5-week visit, 65
completed the 10-week visit, 29 were seen at 15 weeks, 11 at 20 weeks, and one at 25 and 30 weeks after first wearing spectacles. Overall, the mean improvement based on best measured acuity at any visit was 2.9 +/- 1.8 lines. When stratified by baseline visual acuity, those with moderate amblyopia (20/40–20/100) improved 2.9 lines and those with severe amblyopia (20/125–20/250) improved 2.8 lines. Sixty-five of 84 patients (77%) improved at least two lines and 50 of 84 patients (60%) improved at least three lines. If resolution of amblyopia is defined as an interocular difference of one line or less, 27% of patients in this study had resolution with spectacles alone. Better baseline visual acuity and lesser amounts of anisometropia were significant predictors of amblyopia resolution; however, age was unrelated to resolution.

Based on these results, I tend to start patching immediately only if I suspect poor follow-up or if the parents are very anxious to start more aggressive treatment. I will also consider it if a child has a break from school for short period of time and the parents want to do as much treatment as possible during this break. In the absence of these scenarios, I generally treat with spectacles alone until there is resolution of amblyopia or improvement stalls. There are some advantages to this approach. First, some children will not need patching or atropine because glasses alone will result in resolution of amblyopia. Secondly, if patching is needed, the visual acuity is likely to be better when patching is started compared to starting it at the same time as spectacles. A child may be more likely to wear a patch when the amblyopic eye is 20/50 rather than 20/100. Finally, it is ideal to introduce one new treatment at a time, instead of initiating spectacles and patching at the same time. In this way, parents can focus on getting the best compliance with one treatment.

DO WE REALLY KNOW THAT PATCHING WORKS?

The objective of the randomized clinical trial phase of ATS5 was to compare two hours of daily patching combined with one hour of concurrent near visual activities to a control group of spectacles alone if needed for treatment of moderate to severe amblyopia in children age 3–7. PEDIG initiated this study of patching versus control because some authors had questioned whether there was any benefit to patching. Prior to this study, no published study had done all of the following: (1) clearly defined amblyopia at enrollment; (2) incorporated a prolonged spectacles run-in phase with criteria for determining when maximum improvement had occurred; and (3) included a no patching control group. This was a prospective multi-center randomized clinical trial involving 46 PEDIG sites. The main outcome measure was best-corrected amblyopic eye acuity after 5 weeks. If spectacles were needed, optimal refractive correction had to be worn for at least 16 weeks or for two consecutive visits 5 weeks apart without improvement prior to randomization. Only patients with at least two lines of interocular difference (IOD) after completing the spectacles phase of the study were randomized, and three lines of IOD were necessary for inclusion in the primary cohort.

Baseline data were similar between the patching and control groups, including a mean age of slightly over 5 years, proportion with prior to amblyopia treatment of approximately 10%, mean amblyopic eye acuity of approximately 20/80, and mean amblyopic eye spherical equivalent of about 5 D. The primary outcome was the difference between treatment groups in mean logMAR acuity adjusted for baseline acuity after 5 weeks. Overall, this difference was 0.07 logMAR (95% CI = 0.02, 0.12), or slightly less than one line. The dif-
ference between groups was 0.06 (0.01, 0.11) for those with baseline visual acuity of 20/40 to 20/100 and 0.08 (–0.09, 0.25) for those with baseline acuity of 20/125 to 20/400. The patching group had a mean improvement of 1.1 lines compared to 0.5 lines for control.

Since some patients continued to improve after the 5-week primary outcome visit, the best-measured visual acuity at any visit was also determined. Patients wearing a patch improved an average of 2.2 lines compared to 1.3 lines for the control group. The median time to best-measured acuity was 5 weeks in both patching and control groups. However, 38% of those in the patching group had their best measured acuity at 10 weeks compared to only 18% of those in the control group. The treatment effect with patching was not modified by baseline visual acuity, gender, race, age at randomization, prior amblyopia treatments, spectacle phase participation, or cause of amblyopia.

Why did the control group wearing spectacles alone (if needed) continue to improve after apparently stabilizing prior to randomization? First, 5 weeks may have been too short of an interval between visits to detect small amounts of improvement in the spectacles phase of the study. Therefore, some patients may have entered the randomized clinical trial while still improving slightly in spectacles. Second, the visual acuity testing protocol provided measurements in one-line increments, which may not have been sensitive enough to detect slight improvement during the spectacles phase, such as one-half of one line. Third, some children may have just tested poorly at a spectacles phase visit due to lack of effort, fatigue, or test-retest variability, making it appear that they had stopped improving. They would have then been randomized, and if they returned to their baseline level of cooperation at the outcome visit, they would have appeared to have improved between randomization and outcome. It seems unlikely that a learning effect played a significant factor since these patients had performed the test many times prior to the primary outcome visit.

A secondary cohort of patients completed the spectacles phase with only two lines of interocular difference, or three lines of IOD with a 20/32 amblyopic eye and a 20/16 sound eye. Although it was expected that these patients would improve less because there was less room for improvement, the treatment effect relative to control was similar as that of the primary cohort. The patching group improved almost one line with treatment and the control group did not improve. These data reinforce the fact that many patients who reach near-normal visual acuity with spectacles wear will benefit from the addition of patching.

**HOW MANY DAILY HOURS OF PRESCRIBED PATCHING ARE NECESSARY?**

PEDIG randomized patients with amblyopic eye visual acuity between 20/100 and 20/400 to full-time prescribed patching versus 6 hours of daily prescribed patching. Those with moderate amblyopia, defined as 20/40 to 20/80, were randomized to six hours prescribed patching vs. two hours prescribed patching daily. Follow-up visits occurred at 5 and 17 weeks, and the primary outcome was a masked visual acuity assessment 17 weeks after randomization. In those patients with moderate amblyopia, mean visual acuity improvement was 2.4 lines in both the two-hour and six-hour prescribed patching groups, and the mean visual acuity at 4 months was 20/32 in both groups. Those with severe amblyopia also showed no difference between prescribed six hours patching and full-time patching, with both groups improving almost five lines after 17 weeks.

One criticism of this study has been that
many of these patients probably did not wear the patch was much as prescribed. Therefore, the actual treatment received by each group was more similar than the amount of treatment prescribed to each group. This issue highlights the difference between an efficacy and effectiveness study. Studies of efficacy evaluate a treatment under highly controlled conditions. On the other hand, effectiveness studies evaluate a treatment in actual practice, or in the “real world” setting. ATS2 was designed as an effectiveness trial, and, as such, its results are more applicable to clinical practice than an efficacy study. In our practices, we prescribe hours of patching, but we are not able to actually monitor how much the patch is worn once the patient leaves the office.

A second criticism of this study has been that the results are in contrast to the clinical experience of experts. It has been hypothesized that removal of the patch for any period of time results in abnormal binocular interaction that delays visual recovery; consequently, more hours of daily patching would be expected to result in greater improvement. However, clinical experience is not uncommonly at odds with the results of randomized clinical trials. A plausible counterhypothesis is that there is a maximum rate of amblyopia improvement, and that this optimal rate is limited at the biochemical level in the ocular-cortical pathway. It is possible that, for many children, fewer daily hours of patching still provides sufficient stimulus to reach this maximum rate of improvement.

WHAT HAPPENS WHEN PATCHING IS STOPPED?

Amblyopia Treatment Study 2C assessed the risk of recurrence of amblyopia after cessation of treatment. It was a prospective observational study conducted at 30 sites and involving 58 investigators. One-hundred fifty-six children were enrolled who had strabismic or anisometropic amblyopia that had been treated at least three months with at least two hours of daily patching or one drop of weekly atropine and had shown at least three lines of visual acuity improvement. When the investigator felt that they were ready to stop treatment, they were enrolled and followed off treatment. Study visits occurred at 5, 13, 26, and 52 weeks, and recurrence was defined as a two logMAR decrease in acuity that was confirmed by a second visual acuity test. During the one-year follow-up, recurrence occurred in 21% of patients. If we use the alternative definition of a single visual acuity test with a two logMAR reduction in visual acuity (not confirmed with a re-test), the recurrence rate was 24%. Most of the recurrences occurred in the first five weeks after treatment was stopped, with 40% of the 35 recurrences being detected at the 5-week visit. Additional recurrences occurred at 13 weeks, 26 weeks, and 52 weeks, with fewer recurrences during the second half of the year. Interestingly, in those patients who were treated with moderately intense patching, defined at 6–8 hours per day, weaning prior to cessation of treatment seemed to influence whether recurrence occurred or not. Those patients who had no weaning to less intense treatment prior to cessation had recurrence in 11 of 26 cases (42%). This was greater than the 14% recurrence rate (3 of 22) in those patients who weaned prior to stopping treatment.

DOES PATCHING WORK IN OLDER CHILDREN?

Amblyopia Treatment Study 3 enrolled 507 children age 7–18 at 49 sites. Participants had amblyopia associated with strabismus, anisometropia or both. Best corrected visual acuity in the amblyopic eye was 20/40 to 20/400, and best-corrected visual acuity in the sound eye of 20/25 or
better. Patients were randomized to a control group with optical correction alone or to an augmented treatment group, which included optical correction and patching with near activities. In the augmented treatment group, those age 13 or less also used atropine, and those older than 13 did not. The primary analysis compared the proportion of treatment responders, defined as two lines of improvement or more, between the augmented treatment and control groups.

The results showed that treatment with patching and atropine was superior to optical treatment alone in children age 7–12 years. However, there was no significant difference in those children 13–17 years old. There was a significant age effect, as younger children showed a greater effect of augmented treatment relative to control than the older children.

In those children age 7–12 years, augmented treatment was superior to optical treatment regardless of whether prior treatment had been done or not. However, in the older children, only those with no prior treatment had a superior outcome with patching compared to spectacles alone. Those with prior treatment showed no difference between patching and control. Therefore, a trial of patching seems worthwhile in older children and teenagers, particularly if they have had no prior treatment.

DOES ATROPINE WORK AS WELL AS PATCHING?

The first Amblyopia Treatment Study was a randomized clinical trial of patching versus atropine for the primary treatment of moderate amblyopia. Initial treatment was six or more daily hours of patching or daily atropine, and treatment intensity was increased or decreased based on interim visual acuities. Four hundred nineteen children age 3 to nearly 7 were enrolled at 47 sites. Baseline amblyopic eye visual acuity was 20/40 to 20/100. After six months, the patching group had improved 3.16 (2.95–3.37) lines and the atropine group improved to 2.84 (2.61–3.07) lines. The mean treatment group difference was 0.034 logMAR (0.005–0.064). Improvement was initially faster with patching, but atropine was better tolerated by many families based on results of the Amblyopia Treatment Index questionnaire. This study was designed as an equivalence trial—that is, it was designed to show that the difference between treatment groups would be less than a predefined equivalence limit of one line. Although there was a statistically significant difference between treatment groups, these data are a good example of a difference that has statistical significance but is clinically insignificant. Both groups improved about three lines, and most would agree that the difference between groups of 0.3 lines mean improvement is clinically insignificant.

One limitation of this study was that stereoacuity could not be assessed at the six-month primary outcome visit. This problem occurred because many of the patients had used atropine within two weeks of the outcome examination, blurring their vision at near. However, stereoacuity was assessed at the two-year follow-up visit, which was completed by 363 of the 419 children. Treatment from six months to two years was at the discretion of investigators. After two years, the patching group had improved 3.6 lines and the atropine group 3.7 lines, and the mean amblyopic eye acuity was 20/32 in both groups. Sensory outcomes were similar as well, whether including all patients or considering non-strabismic patients only.

HOW OFTEN DOES ATROPINE NEED TO BE USED?

Amblyopia Treatment Study 4 was a randomized trial comparing atropine regimens for the treatment of moderate am-
blyopia in children.\textsuperscript{10} One hundred sixty-eight children age 3 to nearly 7 were enrolled who had baseline amblyopic eye visual acuity of 20/40 to 20/80. Children were randomized to either daily atropine or weekend atropine. After four months, both groups had improved an average of 2.3 lines, and stereoacuity outcomes were similar. This study showed that weekend atropine is as effective as daily atropine for children with moderate amblyopia.

CONCLUSION

In conclusion, several randomized clinical trials conducted by PEDIG have contributed to our understanding of amblyopia treatment. These studies have been designed as effectiveness trials so that the results are more applicable to clinical practice. Several additional PEDIG clinical trials are in progress, including evaluation of the effect of near vision activities while patching, assessment of the effect of augmenting atropine treatment with a plano lens, and comparison of patching to atropine in children age 7 to 13 years.

REFERENCES


\textbf{Key words:} amblyopia, randomized clinical trials, patching, atropine