The Early vs Late Infantile Strabismus Surgery Study: do sources for bias exist in this non-randomised trial?

Kristina Meyer, Heike Breitschwerdt, Gerold H Kolling, Huib J Simonsz, for the Early vs Late Infantile Strabismus Surgery Study Group*

Abstract

**Background**—The Early vs Late Infantile Strabismus Surgery Study Group investigates whether early or late surgery is preferable in infantile convergent strabismus, in a non-randomised, prospective, multicentre clinical trial. The current state of the study after end of recruitment is reported here, focusing on the question of possible sources for bias in this non-randomised trial.

**Method**—The prognostic factors were analysed at baseline in order to check for imbalances between the two treatment groups. Reasons for possible differences are discussed.

**Results**—There is no evidence for clinically relevant inhomogeneities between the two groups concerning the distribution of the three prognostic factors spherical equivalent, degree of amblyopia, and limitation of abduction. The fourth prognostic factor, horizontal angle of squint, differs significantly between the two groups.

**Conclusion**—In the analysis of the final results we may have to account for differences in angle of squint at baseline by its inclusion as a covariate or by stratification.

(Br J Ophthalmol 1998;82:934–938)

Infantile convergent strabismus in most cases starts before the age of 6 months. The child will fail to develop regular binocular vision or lose it rapidly, suppressing the image of one eye. Some children will alternate with either eye, others will use one eye preferentially, resulting in loss of visual acuity in the other eye (amblyopia). Most but not all cases of amblyopia are reversible by patching the healthy eye. Infantile strabismus is very common: its incidence is about 1% in the general population.

Strabismus operations for infantile strabismus can be more precise and secondary motility disorders can be corrected together with correction of the horizontal angle of strabismus in late surgery, resulting in a lower total number of operations. Furthermore, treatment of amblyopia might be more difficult after early surgery, because parents might not recognise the necessity of occlusion therapy after successful surgery.

It is very important for further development in the field of strabismus to know what surgical therapy, at what age, is best for the child. Quite probably different aspects of therapeutic success concerning the number of operations, the degree of binocular vision reached, the final angle of strabismus, and the success in the treatment of the accompanying amblyopia will be affected differently by operating early or late.

This paper reports on the state of the Early vs Late Infantile Strabismus Surgery Study. Our study group is a group of strabismologists and orthoptists from 58 clinics in 11 European countries (see end of paper for list of participating centres). We investigate whether early or late surgery is preferable in infantile convergent strabismus in a non-randomised, prospective, multicentre clinical trial.

Infants between 6 and 18 months of age receive a standardised entry examination. They are operated either before their second birthday in clinics classified as “early”, or between their 32nd and 60th month of age in clinics classified as “late”. All children are evaluated at age 6 in the presence of an investigator representing the opposite treatment group.

After completion of the study, the two groups will be compared regarding degree of binocular vision, angle of strabismus, and visual acuity of the worse eye relative to the better. The analyses scheduled for these three main endpoints are t tests with an overall level of significance of 5%, applying a Bonferroni–Holm correction for multiple testing. Secondary questions of interest

*Members are listed at the end of the paper.

Department of Medical Biometry, University of Heidelberg, Germany
K Meyer, H Breitschwerdt

Department of Ophthalmology, University of Heidelberg, Germany
G H Kolling

Department of Ophthalmology, University Hospital, Rotterdam, Netherlands
H J Simonsz

*Correspondence to: Kristina Meyer, Department of Medical Biometry, University of Heidelberg, Im Neuenheimer Feld 305, D-69120 Heidelberg, Germany.
Accepted for publication 4 March 1998
as number of operations, remaining vertical strabismus, and manifest angle of strabismus during fixation at near will be evaluated in an exploratory way.

Patients were recruited from 1993 to 1996 and the results will be available in 2002. The current state of the study after the end of recruitment is reported here. We will focus on the question of possible sources for bias in this non-randomised trial, mainly discussing comparability at baseline between the two treatment groups.

**Materials and methods**

In early or late surgery for infantile strabismus, the two treatment modalities are extremely dissimilar. Were the trial to be randomised, one would first have to inform the parents of the possibility of surgery at very short notice in case of early surgery, only to postpone surgery for 3 years after randomisation to the “late” group. Even if some very considerate parents would cooperate, the referring ophthalmologists would certainly stop sending patients. Hence, our study group considered a randomised trial not feasible.

One reason for scepticism about non-randomised comparisons is that patient selection may differ between treatments. We therefore have to check carefully for imbalances between treatment groups.

Four potential prognostic factors at baseline have been identified in a pilot study and their degree of homogeneity will be analysed: spherical equivalent, horizontal angle of squint, degree of amblyopia, and limitation of abduction.

The prognostic factors are measured in a standardised way, as described in the study protocol:

(a) The state of refraction is examined with retinoscopy in cyclopia.
(b) The horizontal angle of strabismus is measured with prisms and corneal reflexes during fixation of an object with a light at 50 cm or by estimation of the location of the corneal reflexes during fixation of an object with a light at 50 cm.
(c) The degree of amblyopia is classified into one of the following five categories:
   (1) cross fixation or alternating freely,
   (2) alternating strabismus, but preference of fixation of one eye,
   (3) failure to maintain fixation, but central fixation in fundus in both eyes,
   (4) poor fixation behaviour and eccentric fixation in fundus in one eye,
   (5) poor fixation behaviour and far eccentric fixation in one eye.

(d) The limitation of abduction is measured according to the categories below:
   (1) free abduction, using pursuit movements,
   (2) free abduction, using doll’s head eye movements,
   (3) passing midline but not free, using any method,
   (4) not passing midline, using any method.

Furthermore, we will check the two baseline variables, sex and age at entry examination.

Comparisons of spherical equivalents and age at entry examination will be done performing t tests. Comparisons of angle of squint, degree of amblyopia, and limitation of abduction will be done using Mann–Whitney–Wilcoxon’s U test. Distribution of sex will be compared using a χ² test. In addition, confidence intervals for differences between treatment groups will be reported to facilitate judgment on clinical relevance.

In addition to the four main prognostic factors we compare the occurrence of dissociated vertical divergence (DVD), latent nystagmus, torticollis, vertical deviation in primary position and in left or right gaze, V and A patterns, and the rates of previously prescribed glasses, atropine, or occlusion in the two groups.

**Results**

The recruitment period finished on 31 October 1996. The 58 clinics have entered a total of 532 patients; 231 children have been entered in the early and 301 in the late surgery group.

**SEX AND AGE AT ENTRY EXAMINATION**

The distribution of sex within the two groups is approximately similar. The proportion of male children is 45% in the early group and 51% in the late group, yielding a p value of p=0.18. The 95% confidence interval (CI) for the difference between these proportions is (−3%;15%).

Mean age at entry examination is 11.1 (SD 3.7) months in the group to be operated early and 10.9 (3.7) months in the group to be operated late, corresponding to a p value of p=0.61. The 95% CI for the mean difference in age is (−0.5;0.8) months.

Thus, the differences between age and sex are neither statistically significant nor clinically relevant.

**SPHERICAL EQUIVALENT**

Spherical equivalents are approximately normally distributed in the two groups (Fig 1) with a mean of 2.5 (1.5) D (median 2.25 D) in the early group and 2.2 (1.7) D (median 2.0 D) in the late group, which yields a p value of...
highly significant ($p=0.0001$).

Figure 2 Histogram of horizontal angles of squint at baseline for early and late surgery group. Large angles of squint are observed more frequently in the early group. Mean horizontal angles of squint differ by 2.1 degrees between both groups. This difference is highly significant ($p=0.0001$).

Figure 3 Histogram of degree of amblyopia of the worse eye at baseline for early and late surgery group. There are no statistically significant or clinically relevant differences in degree of amblyopia between the groups.

0.06. The 95% CI for the mean difference between the two groups is ($-0.02;0.57$) D.

Thus, the differences between spherical equivalents are neither statistically significant nor clinically relevant.

HORIZONTAL ANGLE OF SQUINT

Horizontal angles of squint differ between both groups (Fig 2), with a mean of 21.3 (6.1) degrees (median 21.8 degrees) in the early and 19.1 (5.8) degrees (median 20.0 degrees) in the late group. This difference is highly significant with a $p$ value of $<0.0001$. The 95% CI for the mean difference is (1.2;3.2).

In children to be operated early, the angle of squint was measured by estimation of the location of corneal reflexes in $51\%$, with prisms and corneal reflexes in $40\%$ and with prisms and cover test in $9\%$ of all entry examinations. In children to be operated late, the corresponding proportions are $69\%$ for measurement with location of corneal reflexes, $30\%$ for measurement with prisms and corneal reflexes, and $1\%$ (that is, only two children) for measurement with prisms and cover test. This difference in proportions between early and late group is highly significant ($p=0.001$, $\chi^2$ test). The precision of the measurements is lowest for the estimation of the location of corneal reflexes, better for measurement with prisms and corneal reflexes, and best for measurement by prisms and cover test. In children aged 6–18 months, measurement with prisms and cover test can be performed only rarely. The other two methods can be applied independently of the child’s age at entry examination. Note, however, that age at entry examination does not differ between the early and the late group.

In a multivariate regression model we thus correct for the method of measurement (as a dichotomous variable: low $\nu$ high precision—that is, “estimation of location of corneal reflexes” $\nu$ “other methods”, $p=0.01$). Nevertheless, the difference between angles of squint is still significant ($p=0.0003$). What might be the reason for the larger angles of squint among infants to be operated early? Ophthalmologists and parents of children with a large angle of squint might favour an early operation because of clinical or cosmetic considerations. Thus, referral patterns might differ between the groups, leading to the differences that we found. Furthermore, there might occur some sort of measurement bias. Ophthalmologists favouring an early operation might tend to measure larger angles of squint supporting the necessity of an early operation and vice versa. This is why final examinations will be supervised by an ophthalmologist or orthoptist representing the other treatment group (see section “Assessment bias” in the discussion).

We may have to account for this bias in the final analysis of our main end points. This analysis could, for instance, include analyses of covariance with the angle of squint as a covariate.

AMBLYOPIA

The degree of amblyopia is an ordinal variable with the five possible outcomes described above.

The distribution of amblyopia in the two groups is quite similar (Fig 3), with a median class of 2 in both groups and mean 1.9 (0.9) (early group) and mean 2.0 (0.9) (late group). The $p$ value for the difference in distribution between the groups is 0.14. The 95% CI for the mean difference (assuming normality) is ($−0.25;0.06$). The 95% CI for the difference of medians is ($−1.0$).

Thus, there are no statistically significant or clinically relevant differences in degree of amblyopia between the groups.

LIMITATION OF ABDUCTION

Limitation of abduction is an ordinal variable with the four possible outcomes described above.

The distribution of the limitation of abduction is again very similar (Fig 4), with a median of class 2 and mean 2.1 (1.0) in both groups. The $p$ value for the difference in distribution between the groups is 0.75. The 95% CI for the mean difference (assuming normality) is ($−0.14;0.21$). The 95% CI for the difference of medians is (0;0).
Thus, there are no statistically significant or clinically relevant differences in the limitation of abduction between the groups.

OTHER VARIABLES MEASURED AT BASELINE
In addition to the four main prognostic factors, we compare the occurrence of dissociated vertical divergence (DVD), latent nystagmus, torticollis, vertical deviation in primary position and in left or right gaze, V and A patterns, and the rates of previously prescribed glasses, use of atropine, or occlusion between the two groups. According to our pilot study, these variables could not be obtained reliably and reproducibly in children of age 1. Results are summarised in Table 1. Note that, curiously, most variables were found more frequently in the late group, although the average age at examination was equal.

As some variables were optional only, there are quite a few patients with missing information on some of the variables. Proportions of missing data differ between the two groups and the reasons for their absence are unknown to us. We therefore have to be cautious in interpreting the detected differences.

Discussion
In a non-randomised clinical trial we have to be aware of the possible limitations of the validity of our results. Problems might arise because of selection bias, treatment bias, or assessment bias.

TREATMENT BIAS
Amblyopia is treated with occlusion therapy. Atropine and penalisation are allowed only as adjuncts in the treatment of amblyopia, to be determined by the managing ophthalmologist or orthoptist. How to treat amblyopia cannot be prescribed in detail, but this problem is less serious than it may seem, because there is a general consensus that the final goal of treatment of amblyopia is equal and good visual acuity of both eyes. The differences that may result from slightly different treatment goals (for instance, patching less to disrupt fusion to a lesser extent) are small in comparison with the differences caused by lack of compliance of child or parents.

Examination of amblyopia and prescription of glasses have been standardised.

The procedures for surgery are described in the study protocol.19 The names and qualifications of surgeons are documented along with important variables of surgical procedures.

Thus, the treatment bias should be small. Country effects are more likely to occur than biases due to individual centres, since the participating countries vary in health system, medical education, and social background.

Therefore, in addition to the main confirmatory analysis, possible differences between the countries will be explored.

ASSESSMENT BIAS
Unfortunately the two therapeutic modalities, early or late surgery, cannot be blinded. A blinded evaluation of the children at age 6 would be desirable, but this is not possible since the kind of therapy varies with evaluating clinic.

Therefore, instead of blinding, the final examination will be done in the presence of an ophthalmologist or orthoptist from a clinic participating in the study representing the opposite surgery schedule. Thus, assessment bias will be avoided and credibility of the results will be raised.

SELECTION BIAS
After all patients have been entered into the study, we have now presented a final check for imbalances between the treatment groups.

There is no evidence for clinically relevant inhomogeneities between the two groups concerning the distribution of the three prognostic factors spherical equivalent, degree of amblyopia, and limitation of abduction.

Horizontal angles of squint differ significantly between the two groups, which might be due to differing referral patterns.

We might have to account for this difference by modifying our plan for analysis to an analysis of covariance including the angle of squint as a covariate, or by stratification. These possible changes to our study protocol11 will be discussed within the study group.

The remaining examination variables like latent nystagmus and DVD could not be deter-
mined reliably and reproducibly in our pilot study and were, therefore, not taken as prognostic factors. Curiously, the prevalence of most of these was higher in the late group, although the age at examination was equal. It may well be that these prevalences will all rise to equal levels during the future intermediate examinations. We will report on this in future work.

This study is sponsored by grant no Ko 1295 of the DFG (German Research Organisation).

The Early vs Late Infantile Strabismus Surgery Study Group:*