Pooling Data from Similar Randomized Clinical Trials Comparing Latanoprost with Timolol; Medical Results and Statistical Aspects

BY

KATARINA HEDMAN
ABSTRACT

Two different principles were studied. 1st - statistical analysis techniques were used to obtain medical results from a patient population. 2nd - the patient population was used to study the statistical analysis techniques.

Medical conclusions: latanoprost and timolol treatment showed a statistically significant and clinically useful mean IOP-reduction in a typical worldwide clinical trial population. Latanoprost reduced the IOP 1.6 mm Hg more than timolol. The IOP-reduction was maintained with timolol and slightly enforced with latanoprost up to 6 months of treatment. The mean IOP-reduction was maintained during 2 years of latanoprost treatment. The overall risk of withdrawal due to insufficient IOP-reduction with latanoprost was 8%.

The statistical methodological issues are of a general and reoccurring character in trial design of the IOP-reduction: should the statistical hypothesis testing be based on the mean intraocular pressure (IOP) or the proportion of patients who reach a specific IOP level, should the estimate of the IOP or IOP-reduction be based on single eyes, mean of bilaterally eligible and identically treated eyes or the difference between an eye with active treatment and a placebo treated contralateral eye, and is mean of replicated recordings useful? Statistical methodological conclusions: the most effective response variable varies with the selected patient population. Therefore, the trial design process should include a comparison of the variability, test power and required sample size for the possible response variables in a sample of the target population. At minimum a statistical consideration should be done.

Key words: Mean intraocular pressure, target intraocular pressure, latanoprost, timolol, open-angle glaucoma, ocular hypertension, trial design.

Katarina Hedman, Pharmacia Diagnostics, S-751 82, Uppsala, Sweden

© Katarina Hedman 2003

ISSN 0282-7476
ISBN 91-554-5570-0

Printed in Sweden by Universitetstryckeriet, Uppsala 2003
Ophthalmologists raise many questions regarding the measuring and estimation of the intraocular pressure (IOP) in clinical trials. The questions are in general theoretically simple to answer, but difficult for a statistician to communicate in an understandable way. The ophthalmologist might intuitively know that a particular IOP estimate is the most effective one, but he is often unable to bring proof. I have tried to bridge the gap between the two disciplines.
Papers summarized


Contents

1. Introduction................................................................. 6
   1.1 Glaucoma............................................................... 6
   1.2 Pooled data analysis............................................... 6
   1.3 Mean intraocular pressure or target pressure success........ 10
   1.4 Variability of typical mean IOP response variables......... 11
2. Aim.................................................................................. 13
3. Methods and patients..................................................... 14
   3.1 Strategy for pooled data analysis................................. 14
   3.2 Patients........................................................................ 17
4. Results: summary of papers............................................. 22
   4.1 Paper 1.................................................................... 22
   4.2 Paper 2.................................................................... 23
   4.3 Paper 3.................................................................... 26
   4.4 Paper 4.................................................................... 29
   4.5 Paper 5.................................................................... 35
5. Discussion........................................................................... 38
   5.1 Medical concerns of pooled data analysis..................... 38
   5.2 Statistical methodological concerns of pooled data analysis 41
6. Conclusions....................................................................... 44
   6.1 Medical conclusions................................................ 44
   6.2 Statistical methodological conclusions......................... 45
7. Acknowledgements.......................................................... 49
8. References......................................................................... 50
9. Paper 1-5.......................................................................... 54
1. Introduction

1.1 Glaucoma

There are several forms of glaucoma, the most common form being primary open-angle glaucoma. It is an eye disease that is characterized by a progressive damage to the optic nerve with corresponding partial or more extensive loss of the visual field. The prevalence of glaucoma increases with age. Race also affects the prevalence. In a population survey among those over the age of 40 the prevalence was 1.3% and 4.7% respectively in white and black Americans,\textsuperscript{1-2} and it is one of the leading causes of blindness in the elderly.\textsuperscript{3}

There is a strong correlation between intraocular pressure (IOP) and glaucoma. Increased IOP increases the risk of developing glaucoma and reduction of the IOP prevents or reduces the rate of progression.\textsuperscript{4-5} Thus reduction of the intraocular pressure is the accepted treatment of glaucoma, both medical and surgical. Today there are several classes of drugs used as eye drops in the treatment of glaucoma and the present thesis analyzes the effect of two of the most commonly used drugs, the adrenergic beta-receptor blocker timolol and the prostaglandin analog latanoprost.

1.2 Pooled data analysis

Clinical trials of new drugs are performed in several different countries with the purpose of gathering sufficient efficacy and safety information for registration in each specific country or geographical area. Each trial is limited in number of patients, patient character and trial objectives. Therefore, pooling the data of several worldwide trials provides a possibility to obtain a more general estimate of the treatment effect, enables the investigation of prognostic factors, and offers the possibility to explore new objectives. However, new previously unpublished findings from post hoc objectives should be verified by planning for and testing the objectives in a new clinical trial with a new
population. The requirements for pooling the data of different trials are that the trial designs, objectives and case report forms are sufficiently similar and that individual patient data are available. The heterogeneity of the pooled patient population is the key to studying the treatment effect in the various subgroups of interest, whereas, the homogeneity of the measurement settings is the key for pooling the data.

An analysis of pooled data is sometimes referred to as a meta-analysis. However, in a classical meta-analysis only summary data are available, commonly obtained from a literature search. This limits the use of a meta-analysis into constructing an estimate of the summary data that is more general and that can be tested with reinforced power because of the enlarged number of patients. This is an important strategy for objectives that provide small effect differences or where each separate trial has to have a very limited number of patients.

Latanoprost is a phenyl-substituted prostaglandin analogue recently developed for reduction of the IOP in glaucoma or ocular hypertensive patients.6-7 Interestingly, it is a totally new chemical substance in glaucoma treatment and its mechanism of action differs from existing glaucoma medications. It increases the uveoscleral outflow without reducing the aqueuos humor production.8-13 During the development of latanoprost the standard treatment was timolol, a drug that was introduced in the late 1970’s. The latanoprost registration trials used timolol as control treatment. Eight separate registration trials were performed during 1992 to 1999. The trials were conducted in Scandinavia, the UK, USA, Japan, the Philippines, Mexico, China and Korea.14-19 They had similar objectives, design and case record forms and the individual patient data were available. This database gave an extraordinary opportunity to pool the separate trials and extract more information on latanoprost without initiating new trials.
What information was requested? Latanoprost was shown to be about 0.5 to 3.0 mm Hg more effective in reducing the IOP compared to timolol, a result that was statistically significant in all trials but one (the UK).\textsuperscript{14-19} First, the aim was to obtain a general estimate of the treatment effect based on all available trials. Second, to try to find out if the treatment effect differed between specific subgroups of patients (searching for prognostic factors). As potential prognostic factors for IOP-reduction we looked at prognostic factors for developing glaucoma: IOP level at trial initiation, age, ethnic origin, family history of glaucoma and myopia.\textsuperscript{2} The grade of myopia was expressed as the spherical equivalent in this work. Other potential prognostic factors of the IOP-reduction were previous glaucoma treatment, eye color and sex. Patients with previous treatment may not respond sufficiently to single drug treatment. Eye color is an important factor since the melanin in the eye binds to some substances, and thereby causes a diminished treatment effect and sex is interesting from the general health research perspective.\textsuperscript{20} A final aim of the study was to collect information on the long-term effect of latanoprost on IOP.

The three earliest trials (conducted in Scandinavia, the UK and USA) were completed in 1995. They compared latanoprost to timolol during six months of double-masked treatment in a homogenous population, where the majority of the patients were Caucasians who had no previous experience of glaucoma treatment. A pooled data analysis was to be performed to obtain an overall estimate of the IOP-reduction in this type of population. The possibility to search for prognostic factors was limited.

In 1999 all eight registration trials were completed. The measurement setting of the five additional trials differed a little from the three earliest trials, but in total they were all suitable to pool. The most important differences were that the treatment duration was 6 months in the three earliest trials (conducted in Scandinavia, the UK and USA) and 3 months for the other trials, and that the
treatments were open label in the last three trials (conducted in Mexico, Korea and China). The heterogeneous worldwide population gave a broad base for searching for prognostic factors of the IOP-reduction. In addition, the data provided a possibility to summarize the IOP-reduction at different time points during the day, including the peak and trough period of the effect of the two drugs.

The pooled data of all eight trials provided a possibility to study the IOP-reduction after 0.5 and up to 6 months of treatment, a period that is too short to study the long-term effect. Open-angle glaucoma is a progressive disease and requires life-long treatment. Preferably the long-term treatment effect should be studied during several years. Some glaucoma treatment substances, for example timolol, are known to lose some of their initial treatment effect over time. Irrespectively of drug substance a non-negligible number of patients will require a change of substance, multiple substances or surgical interventions over time.

IOP measurements were available during up to 2 years of latanoprost treatment. The Scandinavian and the UK trials had been extended with open label latanoprost treatment up to 18 months after the double-masked 6-month period with latanoprost or timolol treatment. The trial extensions were mainly for adverse event monitoring and the IOP was measured for safety reasons. The different trial clinics had the possibility to choose to participate only in the double-masked period, the double-masked period plus the first trial extension, or the double-masked period plus both trial extensions. Because of this, treatment exposure differed markedly between clinics and initial treatment groups; the trial period of latanoprost treatment was 6, 12, 18 or 24 months. The data of all latanoprost periods were pooled and the IOP-reduction evolution was to be studied stratified for treatment exposure class. In addition, the risk of treatment termination due to insufficient IOP-reduction was studied.
1.3 Mean intraocular pressure or target pressure success

A basic statistical trial design issue in trials of intraocular pressure (IOP) reduction, is whether the primary objective should be based on the patient group mean IOP or the percent patients who have their IOP reduced to or below a specific level (target success).

In individual patients the effect of an IOP reducing treatment is commonly judged by its ability to reduce the IOP down below a specific level (target pressure), with the ambition to prevent further progression of the disease by keeping the IOP below this level. In clinical practice it is recommended that the target is individually determined and continually adjusted. It is generally assumed that a reduction of at least 30% from the IOP at which damage occurred is a reasonable initial target IOP in eyes with a moderate damage of the optic nerve. Higher or lower target pressures may be chosen depending on the degree of optic nerve damage.

In clinical trials the IOP reducing effect is evaluated for a whole group of patients and the shift of the whole distribution of IOP values is of interest and has to be summarized in an appropriate way. This can be done by several different methods and we will demonstrate how some of the most important statistics perform. The strategy of analyzing mean-values of the continuous IOP variable will be compared with the strategy of analyzing target success, using realistic numbers for a typical open-angle glaucoma or ocular hypertension trial population.

Using the target success means dichotomizing a continuous variable. The statistical implication of this has previously been shown in detail; both theoretically based on the Gaussian distribution and empirically based on actual blood pressure distributions in patients with systemic hypertension. Particularly it is well known that proportion differences, ratios and odds-ratios depend not only of the underlying distribution but also on the selected cut-off. It is also well known that the statistical test-power of mean-value based
tests of continuous variables is in general much higher than tests based on
dichotomized variables.30-32

1.4 Variability of typical mean IOP response variables
Ophthalmologists raise many questions regarding the measurement and
estimation of the IOP in specific eyes of individual patients and of the mean
IOP of groups of patients in clinical trials:
Q1. Is it useful with replicated recordings when estimating the IOP in a specific
eye of an individual patient or when estimating the mean IOP of a group of
patients?
Q2. If some of the patients provide IOP values from both eyes and the other
patients provide IOP values from only single eyes, how should the mean
IOP of the group of patients be estimated?
Q3. Is it worthwhile to use the IOP-reduction in adjusted eyes when estimating
the IOP-reduction in a group of patients? That is, using the paired
difference between the IOP of the right and left eye, where one eye is
treated with the active treatment and the contralateral eye is untreated or
placebo-treated.

These questions are all of trial design character that has to be solved before trial
initiation. In addition they have in common their origin in the precision of the
IOP estimate. For trials with parallel treatment group design with the objective
of evaluating the mean IOP-reduction, the response variable is determined by
the mean IOP after treatment or the mean IOP-reduction from baseline (mm Hg
or percent) for each treatment group. The nucleus of these response variables is
the IOP in a single eye, the mean of the IOP in right and left eye or the IOP-
reduction in an adjusted eye. The response variable will of course inherit
variability from the nucleus.

The IOP measurement is a compound of true values and measurement
errors. Sources of variation are patient, eye, replicated recording, observer,
instrument, measurement process and the date and time point of the measurement. Other sources of variation may also be present such as treatment, treatment duration, clinic and prognostic factors. Detailed error sources such as hypofluorescence of the precorneal tear film and accommodation etc have previously been described and will not be covered here.33

A model concept is required to study the impact of the various sources of variation on the IOP measurement. Our simple measurement variability model contains a fixed overall IOP level and three random components: patient, eye and replicate. By using the model we will study how the magnitude of the three random components affects the total variability of the IOP or IOP-reduction in a group of patients for different choices of response variable nucleuses. The theoretical results will thereafter be verified by estimating the variability components and total SD of the various IOP and IOP-reduction response variables in a typical open-angle glaucoma or ocular hypertension trial population.
2. Aim

The purpose of this work was twofold:

1. Studying the effect of a newly developed glaucoma medication in a worldwide population.
   a. Estimating and comparing the IOP after 3 to 6 months of latanoprost or timolol treatment in a worldwide heterogeneous population.
   b. Studying the effect of prognostic factors on the IOP-reduction after treatment with latanoprost or timolol; untreated baseline IOP, previous glaucoma medication, ocular diagnosis, history of glaucoma, eye color, spherical equivalent, sex and age.
   c. Studying the IOP evolution during 2 years of treatment with latanoprost.

2. A methodological study concerning effective trial design.
   a. Comparing the strategy of analyzing mean IOP with the strategy of analyzing the proportion of patients who reach a specific target IOP (target success).
   b. Showing the interrelation of difference in mean IOP, difference in target success and odds-ratio of target success.
   c. Comparing the precision of three different typical IOP reduction response variable nucleuses: the IOP in a single eye, the mean of the IOP in right and left eye and the paired difference of the IOP in right and left eye (adjusted eyes).
3. Methods and patients

3.1 Strategy for pooled data analyses

In general, analysis of variance/covariance technique was used for the estimating and inference testing of the IOP-reduction. Standard errors of the mean and/or 95% confidence intervals accompanied all mean-value estimates. P-values were provided for the majority of estimates. Adjustments for multiple comparisons were not made. The IOP-reduction was further illustrated by descriptive calculations of target success. All data processing was performed with the SAS system.

Latanoprost and timolol showed almost the same clear linear relation of the IOP-reduction and the baseline IOP. This relation was well captured and controlled by using the baseline IOP as a covariate with no covariate interaction in the analysis of the IOP reducing effect of the two drugs (Figure 3A).34

The diurnal IOP-reduction (mean of morning, noon and afternoon) after 6 months of treatment in a homogenous population was analyzed with an analysis of covariance with country, clinic within country, drug and sex as factors, country-by-drug, drug-by-sex and country-by-drug-by-sex as interactions, and untreated IOP as a covariate.

It was considered clinically relevant to study the diurnal IOP and it was also statistically favorable since the standard deviation was smaller for the diurnal IOP than for the IOP at a specific point in time.
The IOP-reduction after 3-6 months of treatment in a heterogeneous population was analyzed with an analysis of covariance with country, drug, sex, diagnosis, with or without previous treatment and age as factors, and untreated IOP as a covariate. The analysis was separately performed for the diurnal IOP-reduction and the IOP at specific time points during the day (morning, noon and afternoon).

The number of patients varied markedly between the different countries and subgroups of interest. Therefore, the interaction effect of the drug and each specific factor was analyzed separately for each factor (by adding the specific drug-by-factor interaction to the initial model). Including all factors and factor interactions simultaneously in the analysis model would give biased factor

---

**Fig. 3** - Scatter plot of changes in diurnal IOP after six months' treatment, compared to untreated diurnal IOP, with a regression line per treatment group.
estimates in the sense that the effect in small and large subgroups would be equally weighted. Also the degrees of freedom would be unnecessary low for the error sum of squares, implying that the model would be over parameterized for its purpose.

The within patient difference in mean diurnal IOP reduction between the eye with the lowest and the eye with the highest spherical equivalence was estimated with an analysis of variance with country and treatment as factors in patients who had different spherical equivalence in the right and left eye.

Despite recommendations that design factors should always be fitted to an analysis model, the analysis of covariance of the heterogeneous population did not include the clinic factor. This was because the wide collection of prognostic factors was important to study, and some subgroups had few observations. In particular, adding a clinic factor would imply a large set of empty subgroups, which may have caused interpretation problems. In addition, information about clinics was not available in the Philippines and Korea trials. The heterogeneous population included at least 91 different clinics and the variation caused by this was assumed to be random. In the analysis model the random residual factor was confounded by the clinic variation and the other factors may have been affected.

The pooled data of the Scandinavian and UK double-masked trials with open label trial extensions gave a homogenous patient population but a very heterogeneous drug exposure. However, the main reason for patients not being followed for 2 years was caused by the trial design and not drug related dropout. The large variation in drug exposure implies that including all patients with their total drug exposure in a single analysis would overestimate the precision of the treatment effect estimate at the later time points of follow-up.

A repeated measurement analysis of variance was used to estimate the mean morning IOP evolution during 0.5 to 2 years of latanoprost treatment, stratified for drug exposure class (6, 12, 18 or 24 months). The difference
between subgroups of patients in mean IOP evolution was analyzed with a split-plot analysis of variance for repeated measurements, stratified for the total population treated for 12, 18 and 24 months. The analyzed subgroups were ocular diagnosis, untreated baseline IOP, with or without previous treatment and treatment exposure class (12, 18 or 24 months).

Time to treatment failure due to insufficient IOP-reduction with latanoprost single drug treatment was estimated with the Life-table method. A Cox proportional hazard regression model was used to compare the risk of treatment failure between subgroups of patients.

The clinic factor was not included in the repeated measurement analysis of variance model of the 2-year data, because within patient changes vary less between different clinics. Neither was the clinic factor included in the Cox proportional hazard regression, because it was desirable to adapt as few factors as possible.

3.2 Patients
All patients included in this work were retrieved from eight clinical registration trials comparing 0.005% latanoprost once daily to 0.5% timolol maleate twice daily in patients with open-angle glaucoma or ocular hypertension (Table 3B). The IOP was recorded in triplicate with a slit-lamp mounted Goldmann applanation tonometer in the morning, noon and afternoon of the baseline visit and last visit. Only morning recordings were obtained at the intermediate visits. The mean of the triplicate recording was used in all calculations if not otherwise stated. The patients and observers were to be well trained and the instrument well calibrated. Each patient was to be followed by a specific observer and instrument. However, the observers and instruments differed between groups of patients.
### TABLE 3

**Percentual (%) Distribution of Baseline Characteristics per Country, Latanoprost (LP) and Timolol (TML) Groups**

<table>
<thead>
<tr>
<th>Country</th>
<th>USA</th>
<th>UK</th>
<th>Scandinavia</th>
<th>Japan</th>
<th>Philippines</th>
<th>Mexico</th>
<th>Korea</th>
<th>China</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of patients</td>
<td>128</td>
<td>140</td>
<td>149</td>
<td>145</td>
<td>183</td>
<td>84</td>
<td>89</td>
<td>95</td>
<td>30</td>
</tr>
<tr>
<td>Male</td>
<td>45</td>
<td>40</td>
<td>66</td>
<td>64</td>
<td>45</td>
<td>40</td>
<td>52</td>
<td>46</td>
<td>57</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18&lt;50</td>
<td>20</td>
<td>14</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>31</td>
<td>25</td>
<td>37</td>
</tr>
<tr>
<td>50&lt;60</td>
<td>26</td>
<td>25</td>
<td>18</td>
<td>22</td>
<td>11</td>
<td>18</td>
<td>28</td>
<td>28</td>
<td>17</td>
</tr>
<tr>
<td>60&lt;70</td>
<td>23</td>
<td>30</td>
<td>39</td>
<td>33</td>
<td>42</td>
<td>37</td>
<td>24</td>
<td>33</td>
<td>17</td>
</tr>
<tr>
<td>70&lt;75</td>
<td>31</td>
<td>31</td>
<td>33</td>
<td>36</td>
<td>39</td>
<td>38</td>
<td>17</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td><strong>Ethnic origin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American/British</td>
<td>21</td>
<td>27</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>73</td>
<td>65</td>
<td>96</td>
<td>98</td>
<td>99</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eye color</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blue/gray/green</td>
<td>8</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>37</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed with brown</td>
<td>63</td>
<td>55</td>
<td>79</td>
<td>83</td>
<td>59</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown</td>
<td>30</td>
<td>35</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Missing</strong></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OH</td>
<td>64</td>
<td>64</td>
<td>56</td>
<td>48</td>
<td>49</td>
<td>43</td>
<td>72</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>POAG</td>
<td>31</td>
<td>32</td>
<td>41</td>
<td>50</td>
<td>34</td>
<td>40</td>
<td>28</td>
<td>40</td>
<td>87</td>
</tr>
<tr>
<td>Exfoliation glaucoma</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>17</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Pigmentary glaucoma</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Family history of OH/glucoma</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>34</td>
<td>37</td>
<td>26</td>
<td>32</td>
<td>31</td>
<td>44</td>
<td>7</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>Missing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Untreated diurnal IOP (mmHg)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-21</td>
<td>14</td>
<td>19</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>17</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>&gt;21-25</td>
<td>52</td>
<td>50</td>
<td>41</td>
<td>48</td>
<td>51</td>
<td>56</td>
<td>67</td>
<td>77</td>
<td>33</td>
</tr>
<tr>
<td>&gt;29-63</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>15</td>
<td>11</td>
<td>17</td>
<td>2</td>
<td>1</td>
<td>43</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>24.4</td>
<td>24.1</td>
<td>25.2</td>
<td>25.5</td>
<td>25.1</td>
<td>24.7</td>
<td>24.1</td>
<td>22.9</td>
<td>29.5</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>3.2</td>
<td>3.6</td>
<td>3.5</td>
<td>3.8</td>
<td>3.5</td>
<td>3.2</td>
<td>2.2</td>
<td>2.2</td>
<td>9.8</td>
</tr>
</tbody>
</table>
Population 1: homogenous population
The data of the Scandinavia, UK and USA trials were pooled.\textsuperscript{14-16} An intention-to-treat population of 829 patients was created (460 on latanoprost and 369 on timolol). For the 55 patients (7\%) who had missing IOP values, the last available value was carried forward and used as a replacement. For 652 patients (79\%) in whom both eyes were identically treated and eligible for trial treatment, the mean of the IOP in the right and left eye was used in the analyses. The IOP was recorded in a double-masked fashion at untreated baseline and after 2 and 6 weeks, 3, 4, 5 and 6 months.

Population 2: heterogenous population
The data of the Scandinavia, UK, USA, Japan, Philippines, Mexico, Korea and China trials were pooled.\textsuperscript{14-19} An intention-to-treat population of 1389 patients was created (737 on latanoprost and 652 on timolol). For the 73 patients (5\%) who had missing IOP values, the last available value was carried forward and used as a replacement. For 1013 patients (73\%) in whom both eyes were identically treated and eligible for trial treatment, in whom both eyes had the same ocular diagnosis, eye color and previous glaucoma treatment, the mean of the IOP in the right and left eye was used. For the 84 patients (6\%) in whom both eyes were identically treated and eligible for trial treatment, but in whom the contralateral eye had a different ocular diagnosis, different eye color or different previous glaucoma treatment, the right eye was used in the analyses.

All 1334 patients who provided IOP recordings after 2 to 8 weeks of treatment were included in the analysis of the initial and late treatment response.

In the Scandinavia, UK, and US trials, the IOP was recorded in a double-masked fashion at untreated baseline and after 2 and 6 weeks, 3, 4, 5 and 6 months. In the Japan trial, the IOP was recorded in a double-masked fashion at untreated baseline and after 2, 4 and 8 weeks and 3 months. In the
other trials, the IOP was recorded in an open label fashion (double-masked in the Philippines) at untreated baseline and after 2 and 6 weeks, and 3 months.

**Population 3: long-term population**

Of the 561 patients that were included in the Scandinavian and UK trials, all 532 patients that provided any IOP values on latanoprost single drug treatment were included in this population. The total latanoprost exposure time was 24 months for 113 patients, 18 months for 103 patients, 12 months for 123 patients, 6 months for 157 patients and less than 6 months for 36 patients. This adds up to a population of 496 patients at 6 months, 339 patients at 12 months, 216 patients at 18 months and 113 patients at 24 months. For 403 patients (76%) in whom both eyes were identically treated and eligible for trial treatment, the mean of the IOP in the right and left eye was used in the analyses.

IOP evolution between 2 weeks and 6, 12, 18 and 24 months was studied in the patients who provided IOP values at those visits, no missing data were replaced; 493 patients at 6 months, 337 patients at 12 months, 215 patients at 18 months and 113 patients at 24 months.

The IOP was recorded at untreated baseline and after 0.5, 2, 3, 4, 6, 6.5, 8, 10, 12, 14, 16, 18, 20, 22 and 24 months of latanoprost single drug treatment for the patients who were initially randomized to latanoprost. The IOP was recorded at untreated baseline and after 2, 4, 6, 8, 10, 12, 14, 16 and 18 months of latanoprost single drug treatment for the patients who were initially randomized to timolol.

**Population 4: for comparing mean strategy with target strategy**

The patient data was retrieved from population 2. All 533 patients who were treated with latanoprost and who had no previous glaucoma medication were included. Thirty-three patients (6%) had missing IOP values. These values were replaced by the last available value carried forward. For the 373 (70%)
patients in whom both eyes were identically treated and eligible for trial
treatment the IOP in the right eye was used.

**Population 5: for comparing precision of mean response variables**
The patient data was retrieved from population 2. All 629 patients with no
previous glaucoma medication, both eyes eligible for trial treatment, both eyes
treated with identical trial treatment, with the same diagnosis and eye color in
both eyes, and who provided triplicate IOP recordings at untreated baseline and
on-treatment were included. Twenty patients (3%) had missing IOP values.
These values were replaced by the last available value carried forward.
4. Results: summary of papers

4.1 Paper 1


Latanoprost reduced diurnal (mean of morning, noon and afternoon assessments) IOP 7.7 (31%, 0.1 SEM) and timolol 6.5 (26%, 0.1 SEM) mm Hg after 6 months’ treatment, from an overall untreated baseline of 24.8 mm Hg (0.1 SEM). Thus the diurnal IOP was reduced 1.2 (18%, 0.2 SEM) mm Hg more with latanoprost than with timolol (p<0.001).

Females reduced diurnal IOP 0.7 (11%) mm Hg less than males for both drugs (p<0.001).

Higher baseline diurnal IOP resulted in larger diurnal IOP reduction during treatment with both drugs. The common linear regression slope was 0.5 (p<0.001).

After 3 and 6 months of treatment, the latanoprost treated patients showed a further decrease in morning IOP of 0.6 (8%, 0.1 SEM, p<0.001) mm Hg from the initial morning IOP reduction obtained at 2 weeks. No such further decrease in IOP was seen with timolol.

In conclusion, latanoprost reduced the mean diurnal IOP more than timolol. The effect observed after 2 weeks was maintained for timolol while latanoprost showed a statistically significant further increase in IOP reduction from 2 weeks to 6 months. Baseline IOP was the only factor of clinical importance shown to be of prognostic value for assessing the IOP reduction.
4.2 Paper 2


Latanoprost or timolol gave statistically significant mean diurnal IOP reduction in the African-American, Asian, Caucasian and Mexican patients, latanoprost with 7.9 mmHg (32%, 0.3 SEM) and timolol with 6.4 mmHg (26%, 0.3 SEM), from an overall untreated baseline of 24.6 mm Hg (0.1 SEM, Figure 1B). This gave a difference of 1.6 mm Hg (25%, 0.2 SEM, p < 0.001, Figure 2B).

![Graph showing diurnal IOP reduction](image)

*Fig. 1.* Diurnal IOP reduction after 3-6 months of treatment with latanoprost or timolol (mean ± SEM, analysis of covariance [ANCOVA]) from overall untreated baseline of 24.6 ± 0.1 mm Hg per country, for patients with or without previous treatment (Tr) and in total. *p* = 0.109, ANCOVA between treatments; *p* < 0.010, ANCOVA between treatments.
The Asian and Mexican patients showed larger differences in mean diurnal IOP reduction between the two drugs (range 1.8–3.1 mmHg) than the European and US patients (range 0.6–1.7 mmHg, p = 0.030, Figure 2B).

Latanoprost produced similar mean diurnal IOP reduction in patients with and without previous glaucoma treatment other than prostaglandins. Timolol treated patients with previous glaucoma treatment reduced mean diurnal IOP 0.9 mm Hg less than timolol treated patients with no previous treatment (13%, 0.3 SEM, p < 0.001, n = 652). The difference in mean diurnal IOP-reduction was 1.2 mm Hg in the previously untreated patients (18%, 0.2 SEM, p < 0.001, n = 963), and 2.3 mm Hg in the patients with previous glaucoma medication (40%, 0.3 SEM, p < 0.001, n = 426).

The latanoprost treated patients diagnosed with primary-open angle glaucoma reduced mean diurnal IOP 0.6 mm Hg (8%, SEM 0.3, p = 0.011, n =
693) more than the latanoprost treated patients diagnosed with ocular hypertension. This was not observed in the timolol treated patients.

Males treated with timolol had a mean diurnal IOP that was reduced 0.7 mm Hg (9%, SEM 0.2, p = 0.004, n = 652) more than females treated with timolol. This was not significant in the latanoprost treated patients where the males reduced mean diurnal IOP only 0.3 mm Hg (4%, SEM 0.2, p = 0.186, n = 737) more than the females.

There was a weak age trend regarding the difference in mean diurnal IOP reduction between the two drugs (p = 0.013, n = 1389). Latanoprost reduced mean diurnal IOP 2.4, 1.9, 1.0 and 1.3 mm Hg more than timolol in the age class of 18-<50, 50-<60, 60-<70 and 70-95 years respectively (SEM 0.4, 0.4, 0.3 and 0.3 respectively).

The mean diurnal IOP reduction was not statistically significantly different in the patients with a family history of OH or glaucoma compared with patients without this history.

The mean diurnal IOP reduction was similar in the eye with the lowest and the eye with the highest spherical equivalence irrespective of treatment. Specifically, the latanoprost treated patients (n = 291) showed no tendency for an increased IOP reduction in the eye with the lowest spherical equivalent compared with the eye with the highest spherical equivalent.

A higher baseline diurnal IOP resulted in a larger diurnal IOP reduction during treatment with both drugs. The common linear regression slope was 0.5 (p<0.001).

The difference in effect between latanoprost and timolol was similar at peak, semi-peak and trough regardless of the patients being previously treated or not.

In total, 26% of the latanoprost-treated patients and 22% of the timolol treated patients showed an increase of the initial treatment response after 2 weeks’ treatment, and reached a better treatment response class after 3 months of treatment (treatment classes were defined as a reduction of <4, 4 - <8, 8 -
<12 or $\geq 12$ mm Hg). On the other hand, 16% of the latanoprost treated patients and 21% of the timolol treated patients experienced a decline in the initial treatment response, resulting in a worse treatment response class after 3 months of treatment. The result was similar after 6 weeks and comparable after 6 months of treatment.

In conclusion, latanoprost or timolol clinically usefully and statistically significantly reduced the mean diurnal IOP in a heterogenous global population. The greatest difference in mean diurnal IOP-reduction between the two drugs was observed in the Asian and Mexican trials.

4.3 Paper 3

For the 113 patients who were treated with latanoprost for 24 months, the mean IOP was 16.9 (SEM 0.3) after 2 weeks of treatment (Figure 4C). After 24 months of treatment the mean IOP was increased with 0.5 mm Hg (0.3 SEM, $p = 0.15$) compared to the 2 weeks level.

For the 102 patients who were treated with latanoprost for 18 months, the mean IOP was 17.4 (SEM 0.2) after 2 weeks of treatment. After 18 months of treatment the mean IOP was increased with 0.2 mm Hg (0.2 SEM, $p = 0.07$) compared to the 2 weeks’ level.

For the 122 patients who were treated with latanoprost for 12 months, the mean IOP was 17.8 (SEM 0.3) after 2 weeks of treatment. After 12 months of treatment the mean IOP was increased with 0.1 mm Hg (0.3 SEM, $p >> 0.05$) compared to the 2-week level.
Figure 4. Morning IOP at untreated baseline and after 0.5 to 24 months of treatment with latanoprost (mean ± 95% confidence interval, n = number of patients).

For the 156 patients who were treated with latanoprost for 6 months, the mean IOP was 18.0 (SEM 0.3) after 2 weeks of treatment. After 6 months of treatment the mean IOP was decreased with 0.7 mm Hg (0.2 SEM, p < 0.01) compared to the 2 weeks level.

The overall rate of withdrawal due to insufficient IOP-reduction as judged by the investigator was 6% (33 out of 532 patients). After adjustment for differences in treatment exposure the overall risk of withdrawal due to insufficient IOP-reduction was estimated to 8%. As expected this risk was closely related to the severity of the ocular disease (Figure 5c).
In conclusion, the initial mean IOP reduction was maintained throughout the 2 years of treatment.
4.4 Paper 4

The difference in mean IOP, the difference in target success and the odds-ratio of target success offer three different statistics to test the difference in IOP between populations. For common glaucoma populations with similar between patient SD, all three statistics test the shift of the IOP distribution (Figure 6D).

![Figure 6. Density function for two different IOP distributions with mean IOP 17 (m_{ref}) and 21 (m_{comp}) mm Hg, respectively, and with an equal SD of 4 mm Hg. The proportion of patients who reach the target 15 mm Hg (T) are denoted for the distribution with mean IOP 21 mm Hg (p_{comp}) as well as the difference in proportion between the distributions (p_{comp}−p_{ref}).](image-url)
Figure 7. Difference in percent of patients \((p_{\text{comp}} - p_{\text{ref}})\) who reach a specific target IOP, “target success”, (13, 15, 17, 19 and 21 mm Hg, respectively) for different competitor populations with lower (13, 14, 16 mm Hg) or higher (18, 19, 20, 21 mm Hg) mean IOP than the reference population with the mean IOP 17 mm Hg. The form of the IOP distributions was identical to the smoothed actual IOP distribution with SD 4 mm Hg. For examples, see paper 4.
Figure 8. Odds-ratio of patients (Odds$_{\text{comp}}$/Odds$_{\text{ref}}$) who reach a specific target IOP, “target success”, (13, 15, 17, 19 and 21 mm Hg) for different competing populations with lower (13, 14, 15, 16 mm Hg) or higher (18, 19, 20, 21 mm Hg) mean IOP than the reference population with the mean IOP 17 mm Hg. The form of the IOP distributions was identical to the smoothed actual IOP distribution with SD 4 mm Hg. For examples, see paper 4.
Therefore, we recommend that only one of the variables be tested for statistical significance and defined as primary endpoint in a clinical trial. This does not preclude one from using the other parameters descriptively as secondary endpoints. Mean IOP differences with corresponding differences in target success are shown in Figure 7D, and mean IOP differences with corresponding odds-ratio of target success are shown in Figure 8D, for typical open-angle glaucoma or ocular hypertensive populations.

Tests based on frequencies of a dichotomous endpoint will in general be less efficient than tests using mean-value statistics (conditional on the appropriateness of the mean-value statistics.\textsuperscript{30-31} Thus the frequency based tests will have lower power and therefore require larger numbers of patients.\textsuperscript{30-31} Nonetheless, the IOP distribution was slightly right-skewed with a low-density tail towards higher IOP values and this deviation from the Gaussian distribution was sufficient to affect the conclusion about the test-power and the required number of patients. For some targets and some changes of the mean IOP it might even be that the Chi-square statistic has equal or better test-power (requiring equal or smaller numbers of patients) than the t-statistic (Figure 9D).\textsuperscript{37} But the closer the IOP distribution is to the Gaussian distribution the larger the gain will be of using the t-statistic regarding the test-power and required number of patients (Figure 10D).\textsuperscript{37} Another advantage of the mean-value based test is that the pooled SD can further be refined by using an appropriate analysis of variance/covariance model.\textsuperscript{34} This method implies a further increase in the test-power and reduction of the required number of patients that outclasses the Chi-square statistic for the IOP variable (Figure 9-10D).\textsuperscript{37}
Figure 9. The number of patients required per treatment group to determine a statistically significant difference in mean IOP between the competitor population and the reference population (range –4 to 4 mm Hg) using a t-test (SD 4 mm Hg) or an analysis of variance/covariance (ancova; thereby refining SD to 3 mm Hg); or to determine a statistically significant difference in the corresponding percentage of patients who reach a specific target IOP, “target success”, (13, 15, 17, 19 and 21 respectively) by using a continuity corrected Chi-square test. The significance level was 5% with a test-power of 80% for a two-tailed test. The Gaussian distribution was assumed for the t-test and ancova, but the smoothed actual IOP distribution was assumed for the Chi-square test. For examples, see paper 4. Note that the curves are only exact for whole mm Hg differences in mean IOP. I recommend that you base your calculations of required number of patients on a table or software that is developed for that purpose.
Figure 10. The number of patients required per treatment group to determine a statistically significant difference in mean IOP between the competitor population and the reference population (range –4 to 4 mm Hg) using a t-test (SD 4 mm Hg) or an analysis of variance/covariance (ancova; thereby refining SD to 3 mm Hg); or to determine a statistically significant difference in the corresponding percentage of patients who reach a specific target IOP, “target success”, (13, 15, 17, 19 and 21 respectively) by using a continuity corrected Chi-square test. The significance level was 5% with a test-power of 80% for a two-tailed test. *The Gaussian distribution was assumed for all tests.* For examples, see paper 4. Note that the curves are only exact for whole mm Hg differences in mean IOP. I recommend that you base your calculations of required number of patients on a table or software that is developed for that purpose.
4.5 Paper 5


When estimating the IOP in a specific eye of an individual patient, using the mean of replicated IOP recordings gainfully decrease the SD of the IOP estimate (Figure 11).

![Figure 11](image1.png)

Figure 11. Percent reduction of the SEM of the mean of independent replicates, per number of replicates (1 - 1/√number of replicates).
When estimating the mean IOP of a group of patients with a high precision instrument, using the mean of replicated IOP recordings or single IOP recordings produces almost the same SEM. However, the lower the precision of the instrument is, the larger is the reduction of the SEM by using mean of replicates (Figure 12).

Figure 12. Percent reduction in required number of patients by using different number of replicates instead of single recording in estimating the patient group mean IOP at a fixed level of the SEM, per quotient of between-replicate variance and the sum of the “other variance”. Where the “other variance” is defined by the total variance minus the between-replicate variance.
Using the mean of the IOP in the right and left eye for bilaterally identically treated patients always gives a smaller SEM of the patient group mean IOP compared to using single eyes. When designing a clinical trial for the evaluation of the IOP-reduction, using adjusted eyes (one eye minus contralateral eye) gives an advantageously lower SEM of the patient group mean value compared to using a design where the patient group mean IOP is based on the mean of the right and left eye for bilaterally identically treated patients, if the sum of the between-eye variance and the variance of mean of replicates is smaller than 2/3 of the between-patient variance. However, adjusted eyes should be avoided if the drug can be expected to have an effect also in the contralateral eye. Decreasing the SD of the IOP gives an advantageous economic consequence, since fewer patients are required to meet a predetermined desired level of the SEM of the patient group mean IOP.

In conclusion, in order to minimize the number of patients required for obtaining a predetermined SEM, the response variable with the lowest SD should be selected.
5. Discussion

A set of patient data can be analyzed for many different purposes. In the present work two different principles were studied. First, statistical analysis techniques were used to obtain medical results from a patient population. Second, the patient population was used to study the statistical analysis techniques. The statistical methodological issues are of a general and reoccurring character in trial design of the IOP-reduction, and they had already been considered briefly in the first phase. Mean of triplicate recordings and mean of bilaterally eligible and identically treated eyes were used whenever applicable. Statistical inference (estimate conduction and hypothesis testing) was based on mean values, and target success was used descriptively (with some few exceptions) to further illustrate the results. The second phase of this work brought proof and clearly illustrated these standpoints.

5.1 Medical concerns of pooled data analysis

A clinical trial is of an experimental nature, deviating more or less from general practice. For example, the compliance to treatment is usually higher limiting the conclusions of the pooled data analyses to patients who show good compliance. However, the enrolled patients were both globally widespread and representative for the majority of open-angle glaucoma or ocular hypertensive patients subjected to medical glaucoma treatment. Therefore, the conclusions should have good generalizability.

The IOP is one of the primary prognostic factors for the development of glaucoma and glaucoma disease progression. A recently published study showed that medical treatment was effective in delaying or preventing the development of glaucoma in ocular hypertensive patients; even with a moderate reduction of the mean IOP of a population (22.5%).4 Another recently published study showed considerable beneficial effects of treatment
that significantly delayed the progress of glaucoma in patients with early glaucoma and moderately elevated IOP.5

Both latanoprost and timolol gave a medically useful mean IOP reduction in all investigated countries. In total, the latanoprost treated patients had their mean diurnal IOP (mean of morning, noon and afternoon) reduced with 32% from untreated baseline and the timolol treated patients with 26%; a 1.6 mm Hg difference.

The difference in mean diurnal IOP-reduction between the two drugs was more pronounced in the Asian and Mexican trials than in the European or US trials, even after adjustment for the “open-label effect” (Figure 2B). This difference cannot be attributed to differences in age, sex, ocular diagnosis, baseline IOP or previous medical therapy, as the analysis was controlled for such differences. Eye color differences could be a possible explanation. A binding of timolol to melanin in the iris has been reported earlier.20 On the other hand, 0.5% timolol twice daily is probably in excess of the concentration required for maximal blockade of the beta receptor.

The subgroup of patients with no previous glaucoma medication yielded a slightly smaller difference between the two drugs in mean diurnal IOP-reduction of 1.2 mm Hg. In the previously treated patients the difference in mean diurnal IOP was 2.3 mm Hg. This was because the patients with previous glaucoma medication (mainly beta-adrenergic antagonists) responded less well to timolol treatment than the patients without previous treatment, a finding that was consistent over the different trials. A selection of non-responders to beta-adrenergic antagonists could have occurred, but this is unlikely. Timolol is known to lose some of its initial effect and some of the patients were treated with beta-adrenergic antagonists for several years.21-22 Therefore, it is possible that the response to timolol in the previously treated patients was less pronounced than it had been initially.

For latanoprost treatment, the IOP-reduction was 0.6 mm Hg larger in open-angle glaucoma patients than in ocular hypertensive patients. For timolol
treatment, the IOP-reduction was 0.7 mm Hg larger in males than in females. These findings are small and of uncertain clinical value, but the trends were detectable in several of the trials. Patients younger than 60 years of age showed a slightly larger difference in mean diurnal IOP-reduction between latanoprost and timolol, in favor of latanoprost, compared to the older patients. This age-trend was considered weak, since it could only be detected in the analysis of the whole pooled data population. Treatment response differences between primary open-angle glaucoma and ocular hypertensive patients and between younger and older patients have not previously been reported with ocular hypertensive drugs. This is probably because the patient populations are seldom large enough to detect such small differences.

Untreated baseline IOP had clinically important and statistically significant impact on the IOP-reduction of both latanoprost and timolol. The higher the baseline IOP is the greater is the IOP-reduction with either of the two drugs, a result that was found in all trials. The phenomenon can in part be explained by “regression towards the mean” that is caused by random variation in the IOP measurement. Because of this it is very important to control differences in baseline IOP when analyzing the IOP reducing effect.

The IOP-reducing effect at peak and trough of different drugs has been brought up to discussion several times, since it is important to have a reliable IOP-reduction during both night and day. This pooled data analysis provided a possibility to study the difference in IOP-reduction between latanoprost and timolol at the time point of peak and trough effects of both drugs in subgroups of the population. Taken into consideration the times for drug application and the times for IOP measurement, the present analysis pointed out that the difference in effect between latanoprost and timolol was similar at peak, semi-peak and trough regardless of the patients being previously treated or not.

The semi long-term IOP-reduction obtained after 3-6 months of treatment was studied in the whole heterogeneous population. However, 6-month data was
only available in the European and US trials. The IOP-reduction observed after 2 weeks was maintained for timolol while latanoprost showed a small but statistically significant further increase in IOP-reduction from 2 weeks to 3-6 months. This could be due to the specific mechanism of action of latanoprost since it changes the extracellular matrix of the ciliary muscle of the eye and thereby increases the uveoscleral outflow.39 This process might take more than 2 weeks and could therefore be a possible explanation for the more pronounced IOP-reduction seen after 2 weeks of treatment.

The long-term IOP-reduction during latanoprost treatment (up to 2 years) was studied in the homogeneous European population of 532 patients, where the majority of patients were Caucasians with no previous glaucoma treatment. The data was obtained from a double-masked 6-month treatment period and two optional trial extensions of 6 and 12 months’ open-label treatment respectively. This caused the treatment exposure to vary markedly between the trial clinics. The subgroup of 113 patients that was followed for 2 years of latanoprost treatment experienced a statistically inconclusive decline in mean IOP-reduction of 0.5 mm Hg after 2 years of treatment compared to the initial IOP-reduction (obtained after 2 weeks of treatment, Figure 4C). Based on this data, it is not possible to determine if this is a true decline in the mean treatment response or a random fluctuation. In total 6% dropped out from single drug latanoprost treatment (33 out of 532 patients) due to insufficient IOP reduction as judged by the investigator. After adjustment for differences in treatment exposure, the overall risk of drop out due to insufficient IOP-reduction was estimated to 8%. As expected, this risk was closely related to the severity of the ocular disease (Figure 5C).

5.2 Statistical methodological concerns of pooled data analysis
When performing a statistical analysis (estimate determination or hypothesis testing) it is very important that the analysis model is well adapted to the actual
object and actual data. This is particularly delicate when analyzing pooled data where the patient character structure often varies markedly between trials, the number of patients in subgroups may vary considerably and the structure of empty subgroups may cause analysis problems. It can be wise to study a selection of possible analysis models before the final model is chosen, to find out if a small but reasonable change in model statement could change the conclusions. If that would be the case, one must investigate closer the implication of the used models and select the one that in the best and most appropriate way estimates the actual objective. This in no way implies that the model that provides the expected or desired values or provides the strongest p-values should be selected – a procedure that is fraudulent and should be condemned.

It is important to be aware of the power of the statistical tests for the objectives of the pooled data. Tests based on pooled data will have higher power than tests based on each separate trial. This implies that small and sometimes clinically unimportant differences in treatment effect between subgroups of patients may be shown to be statistically significant. If such a difference appeared in each of the original trials respectively, although statistically inconclusive in each separate trial, it is highly possible that it is a true difference albeit of uncertain clinical value. Even though the pooled population may give a very high power for tests based on the whole population, the power may still be insufficient for detecting clinically interesting treatment differences in subgroups of patients with few cases. This emphasizes the importance of providing confidence intervals or standard errors of the mean of the treatment effects and not only estimates with p-values.40

The analysis and publication of the three earliest papers followed the medical tradition, in the sense that most estimates were accompanied by p-values. From the statistical point of view it is much more suitable to provide confidence
intervals, in particular when a pooled data analysis is used.\textsuperscript{40} However, the standard error of the mean (SEM) was provided for all estimates; and 2 times the SEM minus the estimate and 2 times the SEM plus the estimate determines an approximate 95\% confidence interval.\textsuperscript{40}

In total, a large set of multiple comparisons was performed without adjusting the p-values or controlling the simultaneous significance level. It is a well-known fact that multiple testing implies an increased risk of obtaining statistical significance purely by chance and not as a result of a true clinical effect.\textsuperscript{41} However, results from post hoc analyses (meta-analyses, pooled data analyses and additional analyses not stated in a trial protocol) should mainly be considered as hypothesis generating or supporting and not as confirming. Therefore, demanding a rigorous control of the simultaneous significance level is probably to be too restrictive for the purpose.

There are recommendations concerning the number of factors that can be fitted to an analysis model for a specific set of data.

The analysis of covariance model of the homogenous population consisted of 829 patients and contained 2 design factors and 3 prognostic factors. The analysis of covariance model of the heterogeneous population consisted of 1389 patients contained 9 prognostic factors. The repeated measurement analysis of variance model of subgroups of the 2-year data consisted of 113 to 493 patients and contained 1 design factor and 0 or 5 prognostic factors. A recommendation for multiple regression analysis (analysis of variance/covariance) is 1 factor per 10 patients.\textsuperscript{42-43}

The total 2-year data consisted of 532 patients, and 33 patients showed insufficient IOP-reduction (33 events, population 3). The Cox proportional hazards model of the total 2-year data contained 3 prognostic factors. A recommendation for survival analysis (Cox regression) is 1 factor per 10 events.\textsuperscript{43-44}
6. Conclusions

6.1 Medical conclusions
The present pooled data analysis comprising eight clinical trials including a total of 1389 open-angle glaucoma or ocular hypertensive patients showed that both latanoprost and timolol treatment statistically significantly and clinically usefully reduced the mean IOP in African-American, Asian, Caucasian and Mexican patients. Latanoprost reduced the mean IOP statistically significantly more than timolol. The difference in mean IOP-reduction was larger in the Asian and Mexican patients than in the European and US patients. For both latanoprost and timolol, higher baseline IOP resulted in a larger IOP-reduction. Latanoprost produced a similar IOP-reduction in patients with or without previous glaucoma medication. Timolol was slightly less effective in the previously treated patients than in the previously untreated patients, but the majority of the previously treated patients had been treated with beta-adrenergic agonists which may account for this observation. The difference in effect between latanoprost and timolol was similar at peak, semi-peak and trough effect of the two drugs regardless of the patients being previously treated or not.

The IOP-reduction observed after 2 weeks was maintained for timolol while latanoprost showed a small but statistically significant further increase in IOP-reduction from 2 weeks to 3-6 months.

A subgroup of 113 patients was followed for 2 years of latanoprost treatment, and this group experienced a statistically inconclusive decline in mean IOP-reduction of 0.5 mm Hg after 2 years of treatment compared to the initial IOP-reduction (obtained after 2 weeks of treatment). Based on this data, it is not possible to determine if this is a true decline in the mean treatment response or a random fluctuation. The overall risk of drop out due to insufficient IOP-reduction was estimated to 8% (adjusted for differences in
treatment exposure). As expected, this risk was closely related to the severity of the ocular disease.

In conclusion, latanoprost and timolol treatment showed a statistically significant and clinically useful mean IOP-reduction in a typical worldwide clinical trial population. The IOP reducing effect was maintained with timolol treatment and slightly increased with latanoprost treatment up to 6 months of treatment. The mean IOP-reduction on latanoprost treatment was maintained during 2 years of treatment, however, this estimate does not include withdrawn patients. The risk of withdrawal due to insufficient IOP reduction increased with the severity of the ocular disease.

6.2 Statistical methodological conclusions
When performing inference testing or calculating an estimate of the IOP or IOP-reduction in open-angle glaucoma or ocular hypertensive patients, the following factors should at the minimum be included in the analysis model (if possible): design factors, baseline IOP and previous treatment. It can be valuable to explore the effect of other factors such as ocular diagnosis, eye color, ethnic origin, sex and age. When designing a clinical trial of the IOP-reduction one should consider the possibility to stratify the randomization for with or without previous treatment, particularly when the number of patients is small. This does not preclude one from also using the previous treatment as a factor in the analysis model or exploring the interaction of previous treatment on the trial treatment.

Commonly the relation of the IOP after treatment or IOP-reduction is linearly related to the baseline IOP (Figure 3A). For this situation it is advantageous to adapt the baseline IOP as a covariate in the analysis model. The interaction effect of treatment and the baseline IOP should always be explored. If the different treatments show sufficiently similar relation of the IOP response variable and the baseline IOP, a simple analysis of covariance
can be used. Otherwise a more complex analysis model must be used or one must consider removing the baseline IOP from the analysis model and/or performing a stratified analysis with respect to the baseline IOP.

Whether or not a variable should be analyzed by a mean-value statistic or by a frequency statistic depends basically on three facts: (1) the distribution of the actual biological variable, (2) the statistical properties of the strategy, and (3) the requirements of the medical authorities. In typical glaucoma or ocular hypertension trial populations, the IOP is a continuous variable with a nearly Gaussian distribution. Virtually all studies of IOP distribution demonstrate a correspondence with a Gaussian bell-shaped curve but with a slight skewness towards higher pressures (Figure 13D).\textsuperscript{45-49} For this type of distribution it is appropriate to use mean-value based tests even with rather small number of patients, and the mean-value strategy will be most efficient from the statistical point of view – particularly if analysis of variance/covariance is used.\textsuperscript{30-31, 50}

If it is required that the IOP reducing effect is evaluated by target success, a target as close as possible to the midpoint between the two IOP distributions (midpoint-target) should be selected in order to maximize the statistical test-power.\textsuperscript{31} Note that if the actual IOP distribution has a higher kurtosis than the Gaussian distribution, the test-power based on a midpoint-target may be close to and sometimes even higher than the test-power based on mean-values. Another advantage of midpoint-targets is that the smaller the between-patient SD is the larger is the difference in target success and the larger is the odds-ratio, for a fixed mean-value difference.

When deciding the minimal difference worthy of detection from a clinical point of view (the alternative hypothesis), it is important to understand the consequence of a mean-value difference between two populations on the individual patients. For clinicians this consequence can be easier to understand if it is expressed as a success proportion or as the number of patients needed to treat to succeed with at least one more patient.
Figure 13. Density function for an actual smoothed IOP distribution (population 4, n = 533 patients) and a Gaussian distribution with the same mean-value and the same SD.

The foundation of statistical consideration is the concept of modeling. A model simplifies a complex real life state into something manageable, preserving the properties of importance for a specific limited situation. Commonly, models are used as analysis tools; e.g. analysis of variance, logistic regression or Cox regression. But the technique is also very useful in controlling measurement variability and precision of estimates by understanding the influence of different sources of variation.

When doing something as ordinary as designing a clinical trial of the mean IOP-reduction, it is valuable to consider the choice of response variable.
Depending on the variability structure of the actual data, using replicated recording, mean of eyes or difference between eyes can gainfully decrease the variability of the patient group mean estimate.

In conclusion, with small and simple statistical efforts in the trial design process, clinical trials can be less costly without affecting the quality of the results. That is, among many possible response variables selecting the one with the best precision or highest test power will reduce the required number of patients. The most effective response variable varies with the selected patient population and therefore, the trial design process should include a comparison of the variability and test power of the possible response variables in a sample of the target population. At minimum a statistical consideration should be done.
7. Acknowledgements

I thank my family and my friends for allowing me to devote valuable time to research.

I thank my supervisors for excellent supervision during my research. Professor Albert Alm, Department of Neuroscience – Ophthalmology, Uppsala University Hospital, Sweden, regarding medical and general issues; Professor Adam Taube, Department of Information Science – Statistics, Uppsala, Sweden, University, regarding statistical and authorial issues. I thank associated professor Johan Harmenberg Medivir, Stockholm, Sweden, for initiating my research within the discipline of medicine; professor Dietrich von Rosen, Swedish Agriculture University, Uppsala, Sweden, for initiating my research within the discipline of statistics; and Dr John Schoenfelder, Global Statistics & Programming, Pharmacia Corporation, High Wycombe, UK, for making it possible for me to perform the current multidisciplinary research.

I thank Pharmacia Corporation for letting me use their clinical data and for financial support. Special thanks goes to the people who work or worked at Pharmacia Corporation at the Marketing Department in Sweden and in the USA and the Statistical and Data Management Department in Sweden and in the UK and the Medical Ophthalmology Department in Sweden.

I thank professor Johan Stjernschantz, Department of Neuroscience – Pharmacology, Uppsala University, Sweden; associated professor Björn Svedberg, Department of Neuroscience – Ophthalmology, Uppsala University Hospital, Sweden; and associated professor Johan Bring, Statisticon, Uppsala, Sweden, for valuable comments on the four initial papers. I thank Professor Thomas Polfeldt, Statistics Sweden, Stockholm, Sweden, for valuable comments on the two papers that dealed with statistical methodological issues; Professor Harry Khamis, Wright State University, Dayton, Ohio, USA, for valuable comments on the first paper that dealed with statistical methodological issues; and Lars Söderström, Pharmacia Diagnostics, Uppsala, Sweden, and Göran Nilsson, Uppsala, Sweden, for valuable information about modeling measurement deviations.

Finally, I thank all my colleges at Pharmacia Corporate and Pharmacia Diagnostics for putting up with having me working part time during my research.
8. References


10. Toris CB, Camras CB, Yablonski ME. Effects of PhXA41, a new prostaglandin F2α analog, on aqueous humor dynamics in human eyes. Ophthalmology 1993;100:1297-1304.


A. Reprinted from European Journal of Ophthalmology; vol. 10, n. 2; authors Hedman K, Alm A; title A pooled data analysis of three randomised, double-masked six-month clinical studies comparing the intraocular pressure reducing effect of latanoprost and timolol; Page No. 98; Copy right 2000; with permission from Wichtig Editore.

B. Reprinted from Survey of Ophthalmology; vol. 47, suppl. 1; authors Hedman K, Larsson L-I; title The effect of latanoprost compared with timolol in African-American, Asian, European, Mexican and US glaucoma or ocular hypertensive patients; Pages No. 80, 82, 83; copy right 2002; with permission from Elsevier Science.

C. Reprinted from Survey of Ophthalmology; vol. 47, suppl. 1; authors Hedman K, Watson P, Alm A; title The effect of latanoprost on intraocular pressure during 2 years of treatment; Pages No. 70, 75; copy right 2002; with permission from Elsevier Science.

D. Reprinted from Journal of Glaucoma; authors Hedman K, Taube A; title Which endpoint: mean intraocular pressure or proportion of successful patients?; copy right 2003; with permission from Lippincott Williams&Wilkins.