When we developed the design of a clinical trial protocol to investigate the clinical relevance of a drug marketed many years ago, calcium dobesilate, for a new indication—erectile dysfunction—we did not have available all the information we needed. Although we started with some encouraging data from in vitro research, no data were available on its clinical effects. Therefore, we did not have information enough to perform a formal sample size calculation for confirmatory statistical analysis because the effect size of the treatment, if there is any, had not been determined. Subsequently, we decided to perform an adaptive interim analysis following the procedure proposed by P. Bauer and K. Köhne based on the Fisher’s product criterion, allowing for a reassessment of the sample size. This article shows the results of such interim analysis and discusses the subsequent decision-making based on this procedure, how to apply the stopping rules, the control of qualitative treatment-stage interactions, and the loss of power to our particular case. The results are critical because they are very close to significance, but formally, the trial has to be continued.

**An alternative approach**

When planning an experiment such as a clinical trial, one must rely on much information that sometimes is not available at the time of protocol writing. Performing a pilot study may be helpful in these situations, but in the medical research setting, to run a sufficiently large pilot study may involve ethical concerns if its results are not to be included in the final inferential statistical procedure—that is, an “external pilot study.” To avoid this, one can take an alternative approach—an “internal pilot study”—that includes an interim analysis and plan for an adjustment of the study protocol on the basis of the results.

We faced this problem to start the clinical development of an old drug in a new indication. Because very encouraging in vitro evidence supporting the possibility of a successful new clinical use for calcium dobesilate was being accumulated, we decided to start development. The chance for clinical irrelevance does exist, however, despite significant results in the in vitro setting. It was not possible, therefore, to actually assume any effect size for formal sample size calculations (as a function of the expected effects and of the variability of the endpoint used to measure them).

We chose an adaptive interim analysis as the most convenient approach: The trial would yield confirmatory results using the minimum sample size that would have been achieved with the best assumptions in a classical fixed one-stage test having suitable information available to calculate it.
Problem statement
Calcium dobesilate, calcium 2,5-dihydroxybenzenosulfonate, is a substance synthesized by Laboratorios Dr. Esteve and marketed since 1969 in a number of countries. A good deal of information is available on its pharmacodynamics, showing that it acts on capillary walls as well as on blood components, and improves or restores the capillary physiology. It significantly reduces endothelial cell desquamation, albumin extravasation, and sorbitol blood levels in diabetic patients; declines platelet hyperaggregation and activation of fibrinolysis, blood hyperviscosity, and erythrocyte rigidity that accompany certain pathological conditions; and improves blood filterability. These properties led to its therapeutic involvement in chronic venous insufficiency, capillary frailty syndromes, diverse retinopathies, and ophthalmic hemorrhages.

Further studies of the effects of calcium dobesilate on vascular reactivity showed that in vitro this substance enhances the endothelium-dependent relaxation. Besides, this effect of calcium dobesilate was inhibited when the isolated tissue was incubated with increasing concentrations of the nitric oxide (NO)-synthase inhibitor N-nitro-L-arginine (L-NNA), while this was reversed with LArg, the substrate in NO synthesis. Most likely, calcium dobesilate acts on the endothelium-dependent relaxing factor NO. In resting endothelial cells, the activity of a constitutive isoenzyme NO-synthase is expressed, which plays a crucial role in the microcirculation homeostasis (blood pressure regulation, inhibition of platelet aggregation and adhesion, and modulation of leukocyte adhesion, which are essential for tissue inflammation). The abnormalities in microcirculation homeostasis as well as the alterations in the reactivity of blood vessels to neurotransmitters and circulating hormones may be underlying the vascular dysfunction induced by diabetes. The in vitro proven support provided by calcium dobesilate to increase the natural constitutive NO production may revert the imbalance that diabetes causes in these processes by restoring the physiologic response of the endothelium cells to neurohormonal control.

The pathophysiology of the erectile dysfunction in diabetics may involve abnormalities of capillary function. As a consequence, drugs with a possible effect in the prevention of pathologies associated with vascular disorders should be the subject of further investigations. Calcium dobesilate has proven to have several positive effects on vascular endothelium; in particular, it may restore the ability of NO synthesis. Because phosphodiesterase inhibitors with proven efficacy in the treatment of erectile dysfunction (that is, sildenafil) increase the amount of NO in vascular cells and thereby cause a vasodilatation, it seems reasonable to investigate the effects of calcium dobesilate in diabetics with erectile dysfunction.

This pool of data encouraged the performance of certain in vitro experiments in resistance penile arteries of the human corpus cavernosum tissues collected from impotent patients at the time of penile prostheses implantation. To study the response of the tissue preparations, the samples were mounted in myographs to evaluate relaxant response after they had been triggered to contract with either norepinephrine or phenylephrine. These experiments demonstrated that calcium dobesilate significantly enhances the endothelium-dependent relaxation evoked by acetylcholine or sodium nitroprusside of previously contracted resistance penile arteries and that this effect can be reverted in presence of L-NNA. On the basis of these summarized results, this drug was considered as an oral therapy for the treatment of erectile dysfunction, especially for diseases affecting the endothelial function.

The next step in the research of this new indication was to start the clinical phase. A proof-of-concept clinical trial had to be designed to establish whether a clinical effect exists. Usually, a pilot study is designed to investigate this and to obtain information for calculating a sample size for a confirmatory test. However, thanks to our approach, we could obviate the former pilot study because the necessary information would be available at the time of the interim analysis.

Statistical plan
We adopted the strategy proposed by P. Bauer and K. Köhne in 1994. The interim analysis allows for a reassessment procedure of the sample size. However, at the final analysis, the common test statistic cannot be used without taking into account the reassessment procedure because increases in the Type 1 error

Nevertheless, this tempting scenario has drawbacks. By observing the treatment differences at planned time points within the course of a clinical trial (interim analyses), the sample size reevaluation is straightforward on the basis of

- the deviation between observed and expected treatment difference.
- an α spending function used to establish critical values for the repeated significance tests.
- conditional power at each step such that at the end of the trial the global unconditional power is preserved.

However, this procedure would greatly inflate Type 1 error. Cui, Hung, and Wang have performed simulation studies to assess the impact of sample size modification according to the previous scheme using the O’Brien-Fleming α spending function, fixing upper limits for sample size increases, coefficients for the conditional power decision-making, as well as an imposed true treatment difference in the sequential design setting. In such simulations, the increase of the sample size led to an increase of the Type 1 error rate, up to 66% in some cases.

While adaptations do interfere with the statistical analysis, an ordered procedure to assume the modifications without compromising the overall significance level or the overall power has been pursued. Research has been undertaken over the last decades to control the Type 1 error rate, and a variety of methods have been introduced for reestimating sample size as well as to allow early termination in the sequential setting. In our work, we employed the method we judged more convenient for our purposes, because its flexibility allows optimizing the risk-benefit ratio in terms of Type 1 and 2 errors and sample size costs for a particular case. We found it also very attractive because it provides ground for more general adaptation of experiments. These characteristics are referred to briefly in the discussion.

This article provides a practical example of the use of an adaptive analysis design to reassess a sample size to deal with the lack of information when planning a Phase 2 clinical trial. We provide a comparison with a classical approach as well.
probability may occur. An ordered procedure for a systematic sample size reassessment and to apply decision rules for early rejection or acceptance of the null hypothesis is possible with the use of some mathematical tools. One of the best known tools is the combination test based on the Fisher’s product criterion. This test relates the two one-sided p-values in the subsamples investigated before and after the adaptation (let \( p_1 \) and \( p_2 \), respectively) with global significance level, \( \alpha \), in a way that \( H_0 \) is rejected at the end of the trial if

\[
p_1 p_2 = \exp \left( -\frac{1}{2} \chi^2 \frac{1 - \alpha}{1 - \alpha} \right)
\]  

[1]

in which

\[
\left( \chi^2 \right)_{\alpha}^2 \left( 1 - \alpha \right)
\]  

[2]

is the 1–\( \alpha \) quantile of the central \( \chi^2 \) distribution with four degrees of freedom. Clearly, if

\[
p_1 \leq \exp \left( -\frac{1}{2} \chi^2 \left( 1 - \alpha \right) \right)
\]  

[3]

the trial can be stopped at the interim analysis with the rejection of \( H_0 \). It can be easily calculated that for an \( \alpha \) level of 0.025, this value is 0.0038.

The pure combination test can be improved, allowing also for early termination in favor of \( H_0 \). If we assume that under \( H_0 \) the observed p-values obtained from stochastically independent samples when a continuous test statistic is applied follow a uniform distribution on [0,1] (if \( H_0 \) is true, this assumption is acceptable under the condition of the stochastical independence of the samples), then a significance lower limit for \( p_1 \) (let \( \alpha_0 \)) can be established so that the interim analysis could lead to the early acceptance of \( H_0 \). It is possible to control the overall Type 1 error probability because it is possible to equate it. When \( \alpha_0 < 1 \), \( p_1 \) could lead to early rejection even if it is higher than \( c_\alpha \) up to a value \( \alpha_1 \) (because the overall \( \alpha \) is kept):

\[
\alpha = \alpha_1 + \int_{\alpha_1}^{\alpha_0} p_2 \, dp_2 \, dp_1
\]  

[4]

This equation allows one to calculate the level of significance \( \alpha_1 \) for an \( \alpha_0 \) chosen. For better understanding, the second term is not more than the sum of all the error probabilities when the trial goes into the second stage (\( p_1 \) is between \( \alpha_1 \) and \( \alpha_0 \), first integration) and is stopped with rejection at the end (\( p_2 \) between 0 and \( c_\alpha /p_1 \), second integration).

One should carefully select a value for \( \alpha_0 \). It is appealing that \( \alpha_1 \) is monotonically increasing as \( \alpha_0 \) decreases; so, accounting for early stopping with no rejection also leads to early rejection of the null hypothesis at substantial levels of \( \alpha_1 \). However, although under the perspective of minimizing the sample size an optimal value for \( \alpha_0 \) could be calculated, it is likely to be lower than 0.25\(^5\) (it is just a minimum–maximum problem). But in such a case, power concerns arise because early acceptance could occur at the end of the first stage even if a trend exists in the good direction. A suitable threshold of \( \alpha_0 \geq 0.5 \) means that in absence of any effect, the procedure will end in half the cases accepting the null hypothesis without proceeding to the second stage, and the level \( \alpha_1 \) is approximately half the global \( \alpha \) level (for \( \alpha = 0.025, \alpha_0 \geq 0.5, \alpha_1 \leq 0.0102 \), calculated from equation [4]). It seems reasonable that just the correct tendency is a requirement for continuation, and therefore the consistency is granted, although in the adapted second stage a large effect is observed (in the case \( p_2 \) is very small, which leads to rejection in the final analysis). In this way, also the possibility of any bias for any qualitative treatment-stage interaction is also controlled. In the Discussion section of this article, rationale is provided for the opposite case (that is, when \( p_1 \) is very small).

The sample size for the second stage now can be calculated by means of a classical fixed sample calculation, in which the significance level is set to the modified value of \( c_\alpha /p_1 \) to comply with the Fisher’s product criterion, considering the estimated variance for the first stage and aiming at a particular conditional power at the end. Note that the Fisher criterion invariably leads to rejection when \( p_2 \) approaches 0.

**Sample data set**

The International Index of Erectile Function (IIEF) was developed by selecting relevant domains of sexual function across various cultures, after the results of tests on patients with erectile dysfunction were reviewed by an international panel of experts. The IIEF questionnaire was examined for sensitivity, specificity, reliability, and construct validity, and the conclusion was made that it is an instrument readily self-administered in research or clinical settings, thereby demonstrating the sensitivity and specificity for detecting treatment-related changes in patients with erectile dysfunction.\(^10\) In addition, we considered that the **efficacy** definition of this pilot study should include identical endpoints that would be required in large Phase 3 clinical trials—which means the evaluation of a meaningful clinical endpoint, in this case, to achieve an erection enabling the patient to penetrate. The IIEF questionnaire provided scores to individual items and five domains as well as a global score.

For a proof-of-concept clinical trial such as this, we decided to use the sum of the scores of three items instead of the global score from the IIEF questionnaire, because these individual scores address specifically the patient’s ability to achieve and to maintain an erection sufficient for intercourse as well as the degree of satisfaction with sexual intercourse, respectively, thus focusing on the clinical effect we were looking for. This concept is more conservative because it would be more difficult to determine differences in just these three items than in the global score when a clinical effect actually is present.

The clinical design of the study comprised a two-week, single-blind placebo run-in period followed by a six-week, double-blind treatment period. Inclusion criteria included the diagnosis of diabetes mellitus irrespective of type and of erectile dysfunction, which was defined as evidence of less than 50% successful attempts to reach and sustain an erection firm enough for intercourse with a partner during a period of at least three months before screening and history of capacity to reach an erection at some time point since the onset of erectile dysfunction, evidenced by nocturnal and morning erections, masturbation, or other sexual activity. Exclusion criteria precluded the recruitment of patients with vascular disorders of the penis tributary from surgical correction, endocrine disorders, penile deformi-
ties, neurological diseases, some concomitant treatments, or in general any serious disease that could interfere with the clinical outcome of the treatment. In this way, a pharmacostimulation test with an intracavernosal injection of prostaglandin E1 and Doppler assessment of the penile vasculature was carried out in all patients during the run-in period to rule out candidates who did not show any response to the intracavernosal injection and therefore are most likely therapy failures that could bias the study results because they cannot be treated with medication. All patients provided a written informed consent before participation, and independent ethics committees approved the study at every participating center.

Patients recruited were administered the IIEF questionnaire to obtain baseline scores and then randomized to receive either calcium dobesilate or placebo in a proportion 1:1 during six weeks. The questionnaire was again administered for the post-baseline score. The primary efficacy endpoint was the change from baseline in the sum of scores of three items. The global as well as the scores of its five domains were considered as secondary endpoints. The null hypothesis tested (which is the same null hypothesis that will be tested after adaptation) is that the change in the primary endpoint under calcium dobesilate from baseline is equal or less than the change under placebo. The adaptive interim analysis was planned once 60 patients were available for analysis. It had been expected that a global sample size of about 260 patients would be necessary for the entire clinical trial. For the interim analysis, a subset of data was locked into the global database of the study after it was considered ready. We initially estimated that a global sample size of about 260 patients would be required for the whole trial to achieve a confirmatory result that supports the decision whether to continue the clinical development of the tested treatment in the new indication. However

Table 1 Effects of the treatments on the scores of the IIEF. Intention to treat population.a

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Pre mean (SD)</th>
<th>Post mean (SD)</th>
<th>Change LSM (SEM)</th>
<th>Change CI</th>
<th>Difference between treatments</th>
<th>p-valueb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo (n = 29)</td>
<td>7.93 (3.2)</td>
<td>8.90 (3.03)</td>
<td>1.09 (0.34)</td>
<td>0.42 – 1.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcium dobesilate (n = 30)</td>
<td>7.27 (2.46)</td>
<td>9.27 (2.85)</td>
<td>2.06 (0.34)</td>
<td>1.37 – 2.74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aTreatments were calcium dobesilate 2 g/day for 6 weeks. Pre: before treatment; post: after treatment; change: adjusted mean (least square mean) of the pre- to post-treatment changes in the scores of each patient; change CI: 95% confidence interval for the adjusted mean response within each treatment group; Diff. CI: 95% confidence interval for the difference in adjusted mean responses between both treatment groups; p-value: one-sided error probability for the rejection of the null hypothesis of less or equal response in the dobesilate group; SD: standard deviation; SEM: standard error of the mean; LSM: least square mean.

bThe critical p-value for rejection at interim analysis is 0.0102, instead of 0.025 (see “Statistical plan” in the text).
ever, we found a strong trend in favor of the active treatment and a low variability in the first stage that allowed minimizing the sample size. Care should be taken to prevent this strong trend leading to misinterpretation of the formal background of the procedure. In spite of a low $p$ value, the null hypothesis cannot be rejected with an overall significance level of 0.025 at this point; that is, the model obliges to continue with a number of patients.

Given the results of the interim analysis, one easily can conclude that if the sample in the first stage had been chosen slightly bigger, then the trial could have been finished with rejection without the need to proceed into the second stage. Moreover, a study with a classical design may be well powered to detect such a clinical difference, and the study may thus conclude with a confirmatory rejection with a sample size (calculated $\alpha$ value that would lead to an early rejection in the Fisher criterion; $n_1$ and $n_2$ the number of patients per group in the first and second subsamples respectively ($n_1 = 30$ in our study). For comparison, the fixed sample size for an optimal test calculated with the estimates from the first subsample is included (parentheses indicate that $N$ is not the sum of $n_1$ and $n_2$ in this case).

The stochastic independence of the test statistics from the different stages of the trial to maintain the model is highly important. This cannot be guaranteed if data are pooled over both stages of the trial. In our particular trial, we took the advantage of the simplest way to ensure stochastic independence because different patients were sampled at each stage. The same endpoint was examined in two random samples; therefore, a design-intrinsic flaw in this sense may be discarded.

The reassessment of the sample size is the simplest of all, and it is the cost of this adaptive procedure in terms of loss of power as far as the sample size considerations not allowing for a correction in the power mirroring that applied for significance level. In our work, we did not include such a correction in the power, because this issue has been investigated also by Bauer and Köhne, who calculated global power of the Fisher's combination test in some scenarios, comparing it with that of the optimal test in the whole sample. The results of such investigations are that the loss of power is very small. It is largest for global power values around 0.6, and is smallest when the two subsamples are equal in size. Following their calculations, for a global power of 0.8, a global significance level of 0.025 and a proportion between the first subsample related to the total (pooled) sample size of 0.7 (this scenario applies in our case, because the proportion of our first subsample is 0.71), the actual power comes to be 0.772, which means a loss of 0.028. This figure also is increased due to allowing for early stopping because there is a chance to accept $H_0$ if $p_1 \geq \alpha_0$ when $p_1 p_2 \leq c_\alpha$ (the gain arising from a rejection $[p_1 \leq \alpha_0]$ that would not have ended so in the combination test without allowing for early stopping $[p_1 p_2 \geq c_\alpha]$ is so small that it can be spurned). This also can be quantified according to the mentioned authors and in our case is lower than 0.0018; therefore, the calculated lowest boundary of the global power of our trial is $0.8 - 0.028 - 0.0018 = 0.7702$. In fact, the mentioned authors did not recommend adapting the pooled sample size because of the loss of power in the adaptive analysis. Of note is that for $\alpha_0$ values less than 0.5, the reduction of power could be more noticeable, because the probability of $p_1 \leq \alpha_0$ and $p_1 p_2 \leq c_\alpha$ becomes higher.

The stochastic independence of the test statistics from the different stages of the trial to maintain the model is highly important. This cannot be guaranteed if data are pooled over both stages of the trial. In our particular trial, we took the advantage of the simplest way to ensure stochastic independence because different patients were sampled at each stage. The same endpoint was examined in two random samples; therefore, a design-intrinsic flaw in this sense may be discarded.

The reassessment of the sample size is the simplest of all, and it is the cost of this adaptive procedure in terms of loss of power as far as the sample size considerations not allowing for a correction in the power mirroring that applied for significance level. In our work, we did not include such a correction in the power, because this issue has been investigated also by Bauer and Köhne, who calculated global power of the Fisher's combination test in some scenarios, comparing it with that of the optimal test in the whole sample. The results of such investigations are that the loss of power is very small. It is largest for global power values around 0.6, and is smallest when the two subsamples are equal in size. Following their calculations, for a global power of 0.8, a global significance level of 0.025 and a proportion between the first subsample related to the total (pooled) sample size of 0.7 (this scenario applies in our case, because the proportion of our first subsample is 0.71), the actual power comes to be 0.772, which means a loss of 0.028. This figure also is increased due to allowing for early stopping because there is a chance to accept $H_0$ if $p_1 \geq \alpha_0$ when $p_1 p_2 \leq c_\alpha$ (the gain arising from a rejection $[p_1 \leq \alpha_0]$ that would not have ended so in the combination test without allowing for early stopping $[p_1 p_2 \geq c_\alpha]$ is so small that it can be spurned). This also can be quantified according to the mentioned authors and in our case is lower than 0.0018; therefore, the calculated lowest boundary of the global power of our trial is $0.8 - 0.028 - 0.0018 = 0.7702$. In fact, the mentioned authors did not recommend adapting the pooled sample size because of the loss of power in the adaptive analysis. Of note is that for $\alpha_0$ values less than 0.5, the reduction of power could be more noticeable, because the probability of $p_1 \leq \alpha_0$ and $p_1 p_2 \leq c_\alpha$ becomes higher.

The stochastic independence of the test statistics from the different stages of the trial to maintain the model is highly important. This cannot be guaranteed if data are pooled over both stages of the trial. In our particular trial, we took the advantage of the simplest way to ensure stochastic independence because different patients were sampled at each stage. The same endpoint was examined in two random samples; therefore, a design-intrinsic flaw in this sense may be discarded.

The reassessment of the sample size is the simplest of all, and it is the cost of this adaptive procedure in terms of loss of power as far as the sample size considerations not allowing for a correction in the power mirroring that applied for significance level. In our work, we did not include such a correction in the power, because this issue has been investigated also by Bauer and Köhne, who calculated global power of the Fisher's combination test in some scenarios, comparing it with that of the optimal test in the whole sample. The results of such investigations are that the loss of power is very small. It is largest for global power values around 0.6, and is smallest when the two subsamples are equal in size. Following their calculations, for a global power of 0.8, a global significance level of 0.025 and a proportion between the first subsample related to the total (pooled) sample size of 0.7 (this scenario applies in our case, because the proportion of our first subsample is 0.71), the actual power comes to be 0.772, which means a loss of 0.028. This figure also is increased due to allowing for early stopping because there is a chance to accept $H_0$ if $p_1 \geq \alpha_0$ when $p_1 p_2 \leq c_\alpha$ (the gain arising from a rejection $[p_1 \leq \alpha_0]$ that would not have ended so in the combination test without allowing for early stopping $[p_1 p_2 \geq c_\alpha]$ is so small that it can be spurned). This also can be quantified according to the mentioned authors and in our case is lower than 0.0018; therefore, the calculated lowest boundary of the global power of our trial is $0.8 - 0.028 - 0.0018 = 0.7702$. In fact, the mentioned authors did not recommend adapting the pooled sample size because of the loss of power in the adaptive analysis. Of note is that for $\alpha_0$ values less than 0.5, the reduction of power could be more noticeable, because the probability of $p_1 \leq \alpha_0$ and $p_1 p_2 \leq c_\alpha$ becomes higher.

Table 2 Reassessment of the sample size,a

<table>
<thead>
<tr>
<th>$p_1$</th>
<th>$a_2$ (points)</th>
<th>$\Delta$ (points)</th>
<th>$\alpha$</th>
<th>$c_\alpha$</th>
<th>Adjusted significance ($c_\alpha/p_1$)</th>
<th>Conditional power aimed at the end (%)</th>
<th>$n_2$</th>
<th>$N = n_1 + n_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0137</td>
<td>1.76</td>
<td>1</td>
<td>0.025</td>
<td>0.0038</td>
<td>0.2774</td>
<td>80</td>
<td>12</td>
<td>42</td>
</tr>
<tr>
<td>0.0137</td>
<td>1.76</td>
<td>0.96</td>
<td>0.025</td>
<td>-</td>
<td>-</td>
<td>90</td>
<td>21</td>
<td>52</td>
</tr>
<tr>
<td>-</td>
<td>1.76</td>
<td>0.96</td>
<td>0.025</td>
<td>-</td>
<td>-</td>
<td>80</td>
<td>-</td>
<td>(54)</td>
</tr>
</tbody>
</table>

a The calculation for the second stage is based in classical fixed sample size calculation using the significance level adjusted by the $p$ value from the first stage and the conditional power aimed at the end. $p_1$ and $a_2$ denotes respectively the one sided $p$ value for the test in the subsample and the estimated standard deviation from the first stage; $\Delta$ the clinical difference worth to detect; $\alpha$ the global significance level; $c_\alpha$ the critical value for rejection in the Fisher’s criterion; $n_1$ and $n_2$ the number of patients per group in the first and second subsamples respectively ($n_1 = 30$ in our study). For comparison, the fixed sample size for an optimal test calculated with the estimates from the first subsample is included (parentheses indicate that $N$ is not the sum of $n_1$ and $n_2$ in this case).
regimens because at the end of the first stage, those not showing relevant effects or excessive toxicity could be removed for the second stage. Thus, the trial could conclude with confirmatory results pertaining to the optimal dose and allowing for a direct eventual switch to Phase 3. A generalization for three stages also has been outlined.

In summary, after the performance of the adaptive interim analysis, we have information about the variability and the size of the treatment effects and a calculated sample size to achieve confirmatory results on the efficacy suitable for a proof-of-concept study could be calculated formally. The differences observed in favor of the tested drug are bigger and the variability lower than expected. Thus the total sample size required to complete the clinical trial is lower than it had been estimated in the planning stage. Actually, the differences found are so large that if we had decided to perform the interim analysis a while later, the study could have been stopped now with rejection. However, since one test for the hypothesis has been performed, as much as 12 patients per group are required additionally to comply with the procedure. In spite of the clear results obtained in the first stage, the possibility of acceptance of the null hypothesis still exists.

**References**


**Acknowledgements**

This clinical trial is being funded by Laboratorios Dr. Esteve, S. A. Av. Mare de Deu de Montserrat, 221, 08041 Barcelona, Spain. The clinical trial is being undertaken within the Czech Republic and the following principal investigators are participating: I. Pavlík (Hospital Krč.–Prague); P. Hanek (Hospital Motol–Prague); J. Vsetíčka and F. Drábek (General Hospital–Jablonec); K. Vlček (General Hospital–Karlovsky Vary), and M. Louda (General Hospital–Hradec Králové). We wish also to thank the support provided by the monitors of the study: C. García (Laboratorios Dr. Esteve, S. A. Barcelona, Spain) and A. Polášková (Verum Staticon s.r.o. Prague, Czech Republic).

**Jesús Villoria,** *MD, is medical writer for Staticon International España, S. A. Parque Empresarial “Cerro de los Gamos,” Bldg. 1, 28224 Pozuelo de Alarcón, Madrid, Spain, +34-91.708.11.15, fax +34-91.708.11.11, email: jvilloria@scr-staticon.com. Pepa García, statistician and head of the Biometrics Department, is also at Staticon International España. Mariano Sust, statistician and head of the Biometrics Dept., and Sebastián Videla, clinical research manager, are at Laboratorios Dr. Esteve, S. A. Av. Mare de Deu de Montserrat, 221 Barcelona, Spain.

*To whom correspondence should be addressed.*