Maximum Likelihood Theory for Retention of Effect Non-inferiority Trials

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

vorgelegt von

Matthias Mielke

aus Uslar

Göttingen 2010

D7

Referent: Prof. Dr. Axel Munk

Koreferent: Prof. Dr. Edgar Brunner

Tag der mündlichen Prüfung: 15.03.2010

CONTENTS

PREFACE

This work deals with the methodical issues of non-inferiority trials whose aim is to demonstrate that a new test treatment is non-inferior to a well-established reference treatment. In the context of this work methodical issues incorporate the development of statistical test procedures and the planning of such trials. The planning of a clinical trial covers the determination of number of test persons required to attain the aspired information content and the optimal allocation of the test persons to the different groups - in the setting of this work a placebo, test and reference treatment group.

The work at hand, basically made up of three papers partially already published, is essentially constituted by the paper Mielke and Munk (2010). In this paper methods for the assessment of non-inferiority in clinical trials are developed for the general setting of arbitrary parametric distributed endpoints, whereas the chronologically preceding paper Mielke et al. (2008) considers the specific setting of censored, exponentially distributed endpoints. The latter paper was essentially influenced by a recent study in the treatment of depression which was kindly placed at the author's disposal by Lilly Deutschland GmbH. However, in order to avoid recurrence of the issues, only the main results of the paper have been presented in Section 5.3. The same applies accordingly to the third paper which considers binary endpoints with the log odds as measure of efficacy (introduced in Section 2.1.2). In contrast to any other setting considered in this work, it has turned out for this particular setting that the asymptotic results for the optimal sample allocation from Mielke and Munk (2010) require further investigations for finite samples and thus need to be adjusted accordingly. From the author's point of view this issue is worth discussing in a separate paper. In this work the issue is examined in Section 6.2.

As already mentioned, all the other sections not explicitly mentioned here trace back to Mielke and Munk (2010).

Acknowledgments

INTRODUCTION

The objective of clinical studies is to provide evidence for the efficacy of a new treatment on the basis of clinical criteria. For example the remission of pain or mortality are feasible criteria to this end. A classical procedure is to prove the superiority of a treatment over placebo and consequently the efficacy of the treatment. However, the inclusion of a placebo group results in ethical concerns for some indications and diseases. Thus, the World Medical Association states in the declaration of Helsinki (WMA, 2000):

> "*The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic, and therapeutic methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic or therapeutic method exists.*"

Therefore, nowadays it is common to compare a new treatment with an established reference (standard) treatment, which often yields smaller differences between the new treatment and the control treatment effect. Consequently, to detect these differences with a given probability, the so-called power of a test decision, a relatively large number of samples is required. This is one reason why *non-inferiority* trials have gained in importance during the last years. In contrast to a superiority trial the aim of a non-inferiority trial is to demonstrate only that the efficacy of a new treatment compared to that of a reference one does not fall below a clinically relevant value, loosely speaking the test treatment is as good as or superior to the reference treatment. For example, a two-arm non-inferiority testing problem is given by

$$H_0 : T \leq R - \Delta \quad \text{vs.} \quad H_1 : T > R - \Delta \tag{1.1}$$

where T and R stand for test (new) and reference treatment, respectively, and $\Delta > 0$ represents a non-inferiority margin which is judged as clinically irrelevant. It is worth noting that $\Delta = 0$ yields a superiority hypothesis. For selective fundamental references, beside the

pioneering work of Dunnett and Gent (1977) we refer to Jones et al. (1996), Senn (1997), Röhmel (1998), D'Agostino (2003) and Munk and Trampisch (2005).

The *ISI Web of Knowledge^{SM}* database demonstrates that non-inferiority becomes a field of vigorous research in the last years. Figure 1.1 presents the distribution of 1490 publications on the topic containing "non-inferiority " or "noninferiority" over the years and its citations (in total 25230).
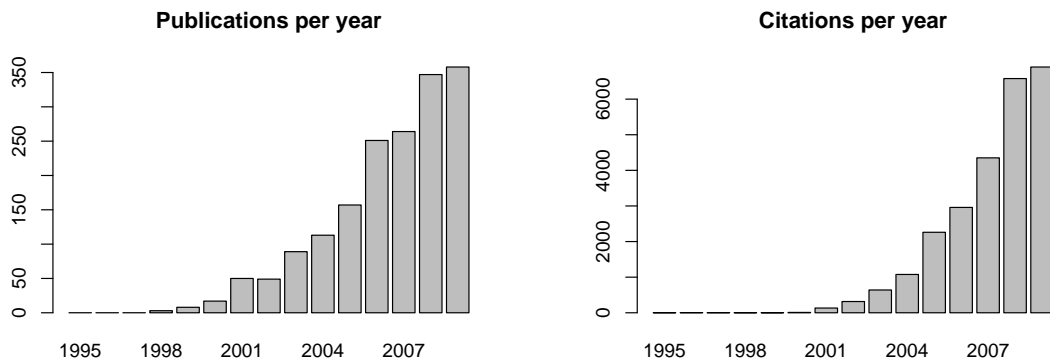
Figure 1.1: *Publications and citations per year on the topic of non-inferiority based on the ISI Web of Knowledge^{SM} database (date: November 2009).*

Beside the fact that a placebo-controlled trial that demonstrates efficacy could be unfeasible due to ethical concerns, a non-inferiority is appropriate when the test treatment is already known for having fewer side effects, improved compliance or leading to a reduction of costs. In this work we focus on this direct comparison of a test and a reference group.

To this end, it is important to note that in active controlled non-inferiority trials without a placebo arm (e.g. the non-inferiority testing problem in (1.1)), the assay sensitivity, i.e. the ability of a study to distinguish between effective, less effective, and ineffective therapies (ICH, 2000), is based on results from historical trials. In this manner the constancy condition is presumed, i.e. the active control effects in the active control trial patient population and the historical trial patient population are assumed to be equal. This assumption is not directly verifiable and its violation could result in statistical uncontrolled errors. Actually, for the treatment of depression there exists evidence that the placebo response is substantial (Dworkin et al., 2005) and that it is increasing over time (Walsh et al., 2002). The above mentioned problems of active controlled trials are discussed by Rothmann et al. (2003) and by Temple and Ellenberg (2000) with regard to the declaration of Helsinki. To this end, the inclusion of a concurrent placebo group is recommended due to the problems of assay sensitivity if the patients are not harmed by deferral of therapy and are fully informed about

alternatives (see e.g. Temple and Ellenberg (2000) and Hypericum Depression Trial Study Group (2004)). Such a design, including a (T)est, (R)eference and (P)lacebo group, has been named by Koch and Röhmel (2004) as the gold standard design. It can be used to demonstrate superiority of either the reference or the test treatment over placebo as well as non-inferiority of the test treatment compared to the reference.

1.1 Retention of effect hypothesis

To demonstrate non-inferiority in the gold standard design we consider the retention of effect type hypothesis

$$H_0 : \theta_T - \theta_P \leq \Delta \cdot (\theta_R - \theta_P)$$

vs. (1.2)

$$H_1 : \theta_T - \theta_P > \Delta \cdot (\theta_R - \theta_P),$$

where $\theta_k \in \Theta \subseteq \mathbb{R}$, $k = T, R, P$, is the parameter of interest, representing the efficacy of a treatment, and $\Delta \in [0, \infty)$ a fixed constant expressing the amount of the active control effect relative to placebo, which should be retained. For a discussion of various issues encountered with the choice of Δ we refer to Lange and Freitag (2005), who provide a systematic review of 332 published non-inferiority studies, and the references given there. Examples for θ_k are (a) $\theta_k = \pi_k$ the success probability of a binary endpoint representing e.g. if the patient achieves remission (Kieser and Friede, 2007), (b) $\theta_k = \lambda_k^{-1}$ with λ_k the expectation of an exponentially distributed endpoint representing e.g. the time until healing or remission (Mielke et al., 2008), (c) $\theta_k = \mu_k$ the expectation of a normally distributed endpoint representing e.g. the FCV (forced vital capacity) in a trial on mildly asthmatic patients (Pigeot et al., 2003). Note that in this set up we presume that large values of θ_k are associated with higher efficacy of the treatment. Compared to absolute hypotheses, e.g. $H_0 : \theta_T \leq \theta_R - \Delta$ with $\Delta > 0$ given in (1.1), the advantage of the hypothesis (1.2) is that it is invariant with respect to rescaling or shifts of the parameters θ_k, i.e. the margin Δ must not be readjusted to the changes of parametrization. Thus, the margin Δ is standardized in that sense and therewith it could easily be compared for different hypothesis and applications, respectively. Further, the margin has the advantage that it has an intuitive and clear interpretation. Rejecting H_0 implies to claim that the test treatment achieves at least $\Delta \cdot 100\%$ of the active control effect, at which both are compared relatively to placebo. Rewriting the alternative in (1.2) as

$$H_1 : \theta_T > \Delta \cdot \theta_R + (1 - \Delta) \cdot \theta_P$$

illustrates that in this case the test treatment effect is greater than a convex combination of the reference and the placebo effect if $0 \leq \Delta \leq 1$. This includes two extremal cases: for $\Delta = 1$
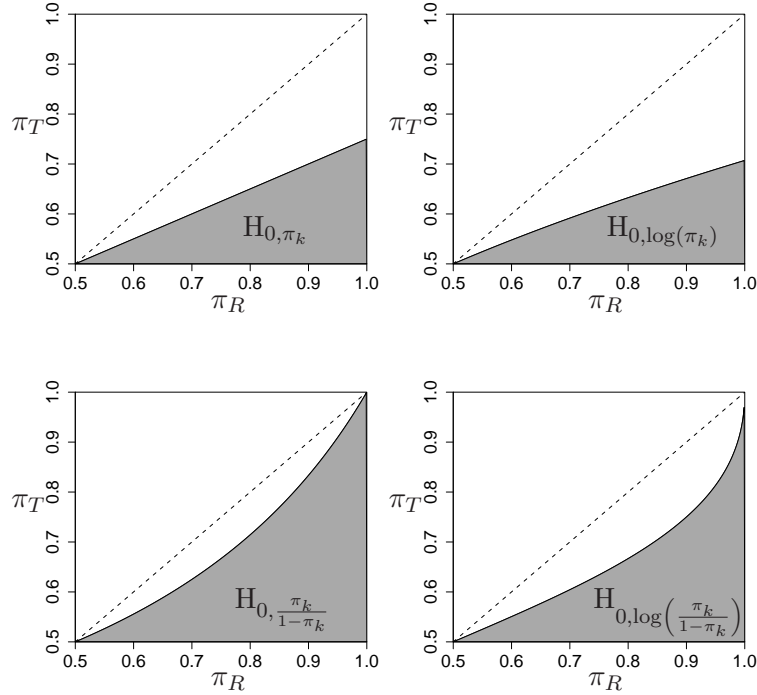
Figure 1.2: *Retention of effect hypothesis $H_{0,\,h(\theta_k)}$ (grey area) for binary endpoints with success probabilities π_k, $\pi_P = 0.5$, $\Delta = 0.5$ and different choices of $h(\theta_k)$.*

we obtain superiority of the test treatment over the reference one (at least $\Delta = 100\%$ of the reference effect is retained) and for $\Delta = 0$ superiority of the test treatment over placebo.

As mentioned above, for binary endpoints a typical choice is $\theta_k = \pi_k$, the success probability. However, in practical application also transformations of the success probability are of interest, e.g. $\log(\pi_k)$, $\pi_k/(1 - \pi_k)$ or $\log(\pi_k/(1 - \pi_k))$. For a comprehensive discussion of several hypotheses for binary endpoints see Röhmel and Mansmann (1999). In order to formalize this we modify the hypothesis (1.2) to

$$H_{0,h(\theta_k)}:\ h(\theta_T) - h(\theta_P)\ \leq\ \Delta \cdot (h(\theta_R) - h(\theta_P))$$

vs.   (1.3)

$$H_{1,h(\theta_k)}:\ h(\theta_T) - h(\theta_P)\ >\ \Delta \cdot (h(\theta_R) - h(\theta_P))$$

where $\theta_k \in \Theta \subseteq \mathbb{R}^d$, $k = T, R, P$, determines the distribution of our endpoints of interest. Here, $h(\cdot)$ is a differentiable, real-valued function on the parameter space Θ measuring the efficiency of a treatment, whereas larger values of $h(\cdot)$ correspond to higher efficiency. In the following, we will omit the alternatives and only state the null hypotheses.

Exemplary for binary endpoints, Figure 1.2 displays the hypothesis $H_{0,h(\theta_k)}$ for different choices of $h(\cdot)$, $h(\pi_k) = \pi_k$, $\log(\pi_k)$, $\pi_k/(1 - \pi_k)$, $\log(\pi_k/(1 - \pi_k))$, for $\Delta = 0.5$ and fixed

$\pi_P = 0.5$. The margin Δ determines the deviation of the hypothesis from the classical superiority hypothesis, represented by $\Delta = 1$ (dashed line). A choice of $\Delta < 1$ results in a non-inferiority hypothesis and a choice of $\Delta > 1$ in demonstrating substantial superiority of T to R.

1.2 Complete test procedure

Typically, the hypothesis (1.2) is embedded in a complete test procedure, where in a first step a pretest for superiority of either the reference or the test treatment over placebo is performed, and in a second step the non-inferiority is investigated via (1.2), confer Pigeot et al. (2003), Koch and Röhmel (2004), Kieser and Friede (2007) and Mielke et al. (2008). All papers figured out as a common result that the pretest is subordinated in the complete test procedure, in terms of that sample size planning can be performed via the non-inferiority test without adjustment to the pretest for superiority. This means the power of the non-inferiority test nearly coincides with the power of the complete test procedure for commonly used alternatives. In addition, the pretest represents a well-investigated testing problem where the parameters of comparison coincide on the boundary of the hypothesis. Thus, in the following we only focus on the non-inferiority hypothesis (1.2) and keep the complete test procedure at the back of mind. Nevertheless, we will touch this issue for the specific setting of censored, exponentially distributed endpoints in Section 5.3.4.

1.3 State of research

Closely related to the retention of effect hypothesis (1.2) is the hypothesis where the treatment effect $\theta_T - \theta_R$ is evaluated relative to a historic active control effect $\tilde{\theta}_R - \tilde{\theta}_P$, which could therefore not been estimated concurrently. For a comprehensive discussion we refer to Holgrem (1999), Hauck and Anderson (1999), Hasselblad and Kong (2001), Rothmann et al. (2003) and Hung et al. (2009). The most problematic issue of such design is the necessity to project the active control effect in the current non-inferiority trial setting (Hung et al., 2009). This issue is not present in the gold standard design, where the active control effect is estimated concurrently.

A nonparametric version of the retention of effect hypothesis (1.2) was already considered by Koch and Tangen (1999). Pigeot et al. (2003) consider (1.2) for normally distributed endpoints. Subsequently, this type of hypothesis was vigorously discussed (see e.g. Hauschke and Pigeot, 2005) and investigated for different types of endpoints. Koch and Röhmel (2004) and Schwartz and Denne (2006) also consider normally distributed endpoints and investigate

(1.2) for θ_k equals the expectation μ_k of the groups $k = T, R, P$, respectively, under homogeneity of variance between the groups. Hasler et al. (2008) extend these results to the case of heterogeneity of the group variances. Mielke et al. (2008) consider censored, exponentially distributed endpoints. Kieser and Friede (2007) investigate binary endpoints with θ_k equals the success probability π_k of each group. In contrast to the normal and exponential case, sample size planning leaves questions open for binary endpoints. In particular, the existing sample size formulas lack in precision, i.e. a deviation between exact and aspired power (cf. Kieser and Friede, 2007). The additional difficulties for binary endpoints arise mainly due to dependency of the variance on the parameters of interest, the success probabilities. In this work we will provide a general approach for general parametric models which allows to close the gaps for binary endpoints as a special case.

1.4   Content and organization

The aim of this work is to provide a general testing methodology based on Wald's maximum likelihood asymptotic to the general retention of effect hypotheses (1.3). This, among others, includes the above mentioned situations as special cases. Furthermore, we obtain tests for Poisson distributed endpoints (for a careful discussion see Section 2.2). Moreover, we discuss the issue of sample size planning and provide in large generality formulas for optimal allocation of samples and accurate approximations for the determination of sample sizes in order to guarantee a certain power. We show that this requires the computation of Kullback-Leibler divergence minimizer in the null hypothesis to an alternative model.

This work is organized as follows. In Section 2, we discuss two clinical trials: first, a trial in the treatment of depressions by investigating whether the patients achieve remission at the treatment end (binary endpoints) measured by the Hamilton rating scale score of depression (HAM-D); second, a study in the treatment of epilepsy by investigating the number of seizures (Poisson distributed endpoints). Beside the "classical" retention of effect hypothesis for binary endpoints, examined by Kieser and Friede (2007), we introduce a retention of effect hypothesis based on the log odds of the success probabilities for binary endpoints, which could be perceived as a "classical" one with a variable non-inferiority margin Δ depending on the success probabilities of the reference and placebo group.

In Section 3, we present the general theory and derive a Wald-type test procedure for the generalized retention of effect hypothesis (1.3), which we denote as *Retention of Effect Wald-type Test (RET)* in the following. In Section 4, we derive sample size formulas and the (asymptotically) optimal allocation for planning a three-armed retention of effect trial. In particular, we include the important case in which the variance is estimated with restriction

to the null hypothesis. This procedure often improves the asymptotic approximation under the null hypothesis (see e.g. Farrington and Manning, 1990; Tang et al., 2007) and therefore is very popular in practice. Our findings can be summarized as follows. The *optimal allocation* when the variance is estimated unrestrictedly turns out to be

$$n_T^* : n_R^* : n_P^* \; = \; 1 : \Delta \, \frac{\sigma_{0,R}}{\sigma_{0,T}} : |1 - \Delta| \, \frac{\sigma_{0,P}}{\sigma_{0,T}} \; , \tag{1.4}$$

where $\sigma_{0,k}$ is the variance within group k, $k = T, R, P$, under the alternative, specified later on in Equation (3.6). Here, n_k^* denotes the number of samples assigned to group $k = T, R, P$. This is shown to be valid in (essentially) any parametric family. Although the asymptotic power will change in general when the variance is estimated restrictedly, we argue that the *optimal allocation* remains unchanged in an asymptotic sense even when the variance is estimated restrictedly to the null hypothesis. However, rather different results occur for the *total sample size* determination, depending on the variance estimation method, restricted or unrestricted. As the optimal allocation (1.4) depends on the choice of the alternative we show in Section 4.1.3 that one may use the allocation $1 : \Delta : (1 - \Delta)$ as a rule of thumb. We show that this allocation is more appropriate in terms of power than the commonly used allocation 2:2:1, as well as the balanced allocation, if $\sigma_{0,P}^2 / \sigma_{0,T}^2$ is (roughly) less than 2. It is important to note that this result is valid in great generality, independent of the distribution of the endpoints and of the formulation of the hypothesis (1.3).

In Section 5, we will revisit our examples introduced in Section 2 to demonstrate and to discuss the results of the previous sections in detail. In addition, we discuss censored, exponentially distributed endpoints by means of summarizing the results from Mielke et al. (2008). We show that sample size reductions and therewith reductions in the costs of a trial with up to 20% and more are possible by reallocating to the optimal allocation instead of a balanced or the commonly used 2:2:1 allocation. In particular, it turns out that our sample size formula for binary endpoints significantly improves the precision of the existing one by Kieser and Friede (2007) in terms of that the exact power is close to the aspired one.

In Section 6, we briefly discuss the finite sample behavior of the normal approximation of the presented procedures and those of the asymptotically optimal sample allocation in detail. As a major point we observe that the asymptotically optimal sample allocation does not provide satisfactory finite sample approximations for the log odds retention of effect hypothesis, in particular, when the success probabilities of the test and reference group become large and the variance is estimated with restriction to the null hypothesis. According to this observation we determine numerically the regions of parameters in the alternative where the allocation $1 : \Delta : |1 - \Delta|$ works quite well and thus no adjustment of the optimal allocation is necessary. In settings, where the allocation needs to be adjusted, we recommend to numerically determine

the finite sample optimal allocation. Nevertheless, for the other settings considered in this work, where it is appropriate to estimate the variance restricted to the null hypothesis, namely binary endpoints with "classical" hypothesis and Poisson endpoints, it turns out that the asymptotically optimal allocation provides good finite sample approximations.

In Section 7, we briefly comment on R-software for analysis and planning of the RET, which is provided in the appendix, in order to allow the reader to reproduce the presented results and to make the presented methodology directly applicable. Finally, we conclude with a discussion in Section 8.

RETENTION OF EFFECT MODELS AND EXAMPLES

In this section we introduce two clinical non-inferiority trials, one in the treatment of epilepsy and the other one in the treatment of depression, and we define retention of effect hypotheses, which are of interest within these examples.

2.1 Binary endpoints: Treatment of depression

Binomial or binary endpoints, respectively, are most commonly used in non-inferiority trials (Lange and Freitag, 2005). In this section we introduce a clinical trial in the treatment of depression from Goldstein et al. (2004) which was also used by Kieser and Friede (2007) for illustration. In particular, we will find different answers concerning the planning of this study (see Section 5.1). This randomized, double-blind trial compares duloxetine (**T**est treatment) to paroxetine (**R**eference treatment) and **P**lacebo with regard to efficacy and safety. In the therapy of depression, achieving remission is the clinically desired goal (Nierenberg and Wright (1999)), whereas remission is defined as maintaining the Hamilton rating scale score of depression (HAM-D) total score at ≤ 7. Table 2.1 presents for each group, $k = T, R, P$, the total numbers of patients and the fractions of patients, who achieved remission at week 8 (end of treatment).

2.1.1 "Classical" retention of effect hypothesis

Following Kieser and Friede (2007), we consider the retention of effect hypothesis with $h(\pi_k) = \pi_k$ for demonstrating that duloxetine is non-inferior over the reference treatment paroxetine,

$$H_{0,\pi_k}: \quad \pi_T - \pi_P \ \leq \ \Delta \cdot (\pi_R - \pi_P) \tag{2.1}$$

where π_k represents the remission probability of treatment $k = T, R, P$ at the end of treatment. The hypothesis is presented in Figure 1.2 (upper left) for $\Delta = 0.5$,

Table 2.1: *Three-armed clinical trial in treatment of depression*

Treatment	No. of patients	No. of Patients achieved remission	Fraction of patients achieved remission
Placebo	88	26	29.55%
Reference	84	31	36.90%
Test	86	43	50.00%

2.1.2   Log odds retention of effect hypothesis

The retention of effect hypothesis for binary endpoints so far considered in the literature and introduced above (2.1) tolerates a large reduction of the success probability of the test treatment π_T if π_R is large or close to one, respectively. Depending on the application this might be an undesired effect. To overcome this drawback Kim and Xue (2004) and Zhang (2006) discuss non-inferiority testing with a variable margin. Zhang points out, that for anti-infective trials a treatment difference of 15% in the response rates is often accepted for a control rate of 75%, however it may be unacceptably large for a control rate of 95%. Therefore, we consider the evaluation of these trials on the basis of the log odds, which also is the natural parametrization for binary endpoints, yielding the hypothesis

$$H_{0,\log(\frac{\pi_k}{1-\pi_k})}: \ \log\left(\frac{\pi_T}{1-\pi_T}\right) - \log\left(\frac{\pi_P}{1-\pi_P}\right) \ \leq \ \Delta \cdot \left(\log\left(\frac{\pi_R}{1-\pi_R}\right) - \log\left(\frac{\pi_P}{1-\pi_P}\right)\right) \ (2.2)$$

presented in Figure 1.2 (lower right). We argue that this overcomes the difficulty of large reduction of the success probability of the test treatment. For π_R close to π_P the hypothesis keeps nearly unchanged compared to the hypothesis (2.1). However, for increasing π_R the success probability of the test treatment π_T is forced to be closer to the one of the reference treatment π_R and even nearly the same for π_R very close to one. As usual, the margin Δ determines the deviation of the hypothesis from the classical superiority hypothesis, represented by $\Delta = 1$. A choice of $\Delta < 1$ results a non-inferiority hypothesis and a choice of $\Delta > 1$ in demonstrating substantial superiority of T over R. Figure 2.1 displays for illustration the hypothesis (2.2) for different choices of Δ and $\pi_P = 0.5$.

The logodds retention of effect hypothesis (2.2) can be perceived as the "standard" retention of effect hypothesis (2.1), whereat the non-inferiority margin Δ depends on the success probabilities π_R and π_P of the reference and the placebo group, respectively, i.e. the margin Δ represents in that sense a variable margin. This is due to the fact, that the hypothesis (2.2) with margin $\Delta_{logodds}$ coincides with the hypothesis (2.1) when the margin for (2.1) is chosen
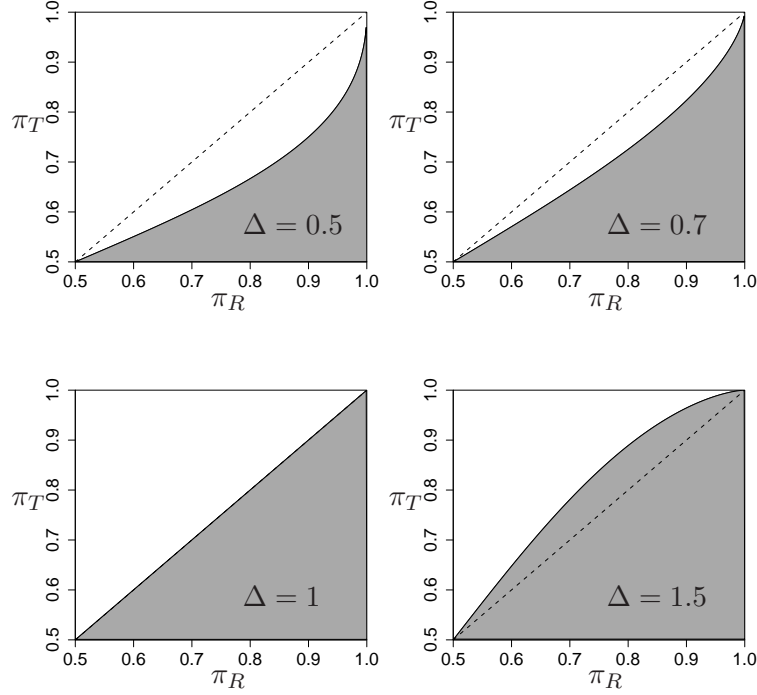
Figure 2.1: *Log odds retention of effect hypothesis $H_{0,\log\left(\frac{\pi_k}{1-\pi_k}\right)}$ (grey area) for binary endpoints with success probabilities π_k, $\pi_P = 0.5$ and different choices of Δ.*

according to

$$\Delta(\pi_R, \pi_P, \Delta_{logodds}) = \frac{\frac{\beta}{1+\beta} - \pi_P}{\pi_R - \pi_P} \tag{2.3}$$

with

$$\beta = \exp\left(\Delta_{logodds} \cdot \log\left(\frac{\pi_R}{1-\pi_R}\right) + (1 - \Delta_{logodds}) \cdot \log\left(\frac{\pi_P}{1-\pi_P}\right)\right).$$

For illustration we display this relation in Figure 2.2 for $\Delta_{logodds} = 0.5$, i.e. how the margin $\Delta_{logodds}$ from hypothesis (3.2) carries over to Δ from (2.3) in dependency of π_R and π_P. As already mentioned above, Δ converges to $\Delta_{logodds}$ as π_R tends to π_P and is equal one for $\pi_R = 1$. Moreover, we obtain that $\Delta(\pi_R, \pi_P, \Delta_{logodds}) \geq \Delta_{logodds}$ for $\pi_R \geq \max\{0.5, 1 - \pi_P\}$, thus, throughout for $\pi_P \geq 0.5$. For $\pi_P \leq 0.5$ we obtain that $\Delta(\pi_R, \pi_P, \Delta_{logodds}) \leq \Delta_{logodds}$ for $\pi_P \leq \pi_R \leq 1 - \pi_P$, whereat $\Delta(\pi_R, \pi_P, \Delta_{logodds})$ is minimal for $\pi_R = 0.5$.

2.2 Poisson endpoints: Treatment of epilepsy

Typical examples of Poisson distributed endpoints can be found for example in the treatment of angina pectoris, nausea and epilepsy, see Layard and Arvesen (1978), where the number of
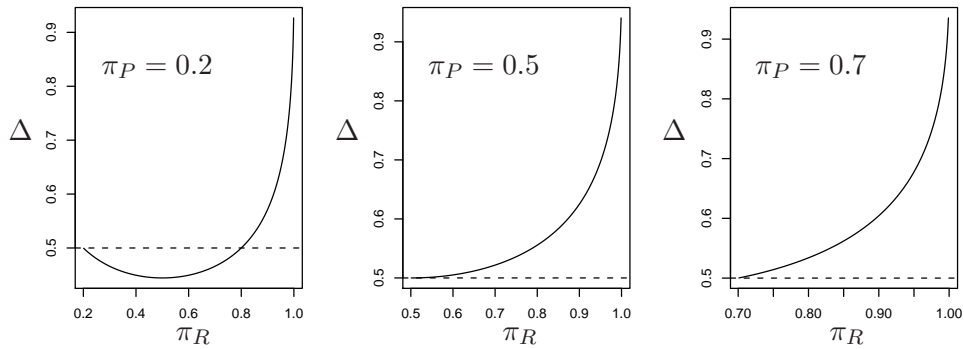
Figure 2.2: *Dependency of the margin Δ from (2.3) on π_R and π_P for $\Delta_{logodds} = 0.5$.*

attacks are counted within a specified time interval, or in the treatment of depressions where the (waiting) time until healing or remission is observed (see e.g. Mielke et al., 2008). Here, we reconsider the randomized, double blind cross-over trial in the treatment of epilepsy from Sander et al. (1990) which compares a new treatment (lamotrigine) as an add-on treatment to a placebo add-on by means of 18 patients. Table 2.2 presents the total number of seizures within the treatment weeks 9-12. Note that Mohanraj and Brodie (2003) highlight that for evaluating anti-epileptic drugs (AED) as add-on treatment the standard endpoint is the manipulation in the number of seizures.

Table 2.2: *Three-armed clinical trial in treatment of epilepsy*

Treatment	No. of Patients	Total no. of seizures	Mean no. of seizures per patient
Placebo add-on	18	338	18.78
Reference add-on	18	295	16.39
Test add-on	18	288	16.00

As AED trials performed in the past are two-armed, either placebo- or active-controlled (for an overview see Mohanraj and Brodie, 2003), we add an artificial reference treatment group (also displayed in Table 2.2) with equal size of 18 patients and seizures of same order of magnitude as seizures under the test treatment for illustration purposes of our procedures.

We presume that the number of seizures of each patient follows a Poisson distribution determined by the group affiliation (T,R,P), i.e. the observations are from $X_{k1}, \ldots, X_{kn_k} \overset{i.i.d.}{\sim} Pois(\lambda_k)$ for $k = T, R, P$ with $n_P = n_T = n_R = 18$. Table 2.2 presents the total number of

seizures in each group, $X_k = \sum_{i=1}^{n_k} X_{ki}$, $k = T, R, P$. As in this setting small values of λ_k, representing less seizures, are desired we choose $h(\lambda_k) = -\lambda_k$ which yields the retention of effect hypothesis

$$H_{0,-\lambda_k} : \ \lambda_P - \lambda_T \ \leq \ \Delta \cdot (\lambda_P - \lambda_R) \tag{2.4}$$

for demonstrating that the test treatment is non-inferior over the reference one.

We would like to mention that Lui (2005) discuss the assessment of non-inferiority for a new treatment over a standard treatment under Poisson distributed endpoints, i.e. the consider a two-arm trial without placebo group. To this end, Lui compares a procedure using the normal approximation and a exact test procedure, whereat he figures out that approximative procedure works quite good in general and that exact procedure does not provide significant improvements.

2.3 Further models

Table 2.3: *Survey of retention of effect hypotheses*

Distribution	θ_k	$h(\theta_k)$	σ_k^2
Normal (Pigeot et al., 2003)	(μ_k, τ^2)	μ_k	τ^2
Normal (Hasler et al., 2008)	(μ_k, τ_k^2)	μ_k	τ_k^2
Binary (Kieser & Friede, 2007, this work)	π_k	π_k	$\pi_k(1 - \pi_k)$
Binary (Figure 1.2, lower right)	π_k	$\log(\pi_k/(1 - \pi_k))$	$(\pi_k(1 - \pi_k))^{-1}$
Exponential (Mielke et al., 2008)	λ_k	$\log \lambda_k$	1
Poisson (this work)	λ_k	$-\lambda_k$	λ_k
Gamma	(α, β_k)	$\alpha \cdot \beta_k \ [= \mathrm{E}X_k]$	$\beta_k^2 \alpha^{-1}$
Weibull	(λ_k, β)	$\lambda_k \ [= \mathrm{E}X_k \cdot (\Gamma(1 + \beta^{-1}))^{-1}]$	$I_{11}(\lambda_k, \beta)$

In Table 2.3 we summarize various endpoints together with some common retention of effect hypotheses. Moreover, we have included some models which have not been used in the context of retention of effect hypothesis, including the Weibull- and Gamma-family. However, these endpoints are of practical interest as recent non-inferiority trials by Yakhno et al. (2006) and Gurm et al. (2008) highlight. We will not discuss all these situations in detail, but we mention that our methodology immediately applies to these situations.

THEORY: WALD-TYPE TEST

In this section we derive a Wald-type test procedure for the generalized retention of effect hypothesis (1.3) introduced in Section 1 and discuss the estimation of the variance with restriction to the null hypothesis. This generalizes and unifies specific results for binary, normally and exponentially distributed endpoints. Based on these results, we provide the theory for sample size planning in the next Section 4.

Model assumptions: Let X_{ki} for $i = 1, \ldots, n_k$ be independently distributed according to a parametric family of distributions with densities $\{f(\theta, \cdot) : \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^d$, and parameters $\theta_k \in \Theta$, $k = T, R, P$, where T, R and P stand for test, reference and placebo group, respectively. We presume that the family of probability densities $\{f(\theta, \cdot) : \theta \in \Theta\}$ is sufficiently regular to obtain asymptotic normality of the ML-estimators (MLE) of the parameter θ with non-singular covariance or Fisher-information matrix, respectively, e.g. an exponential family or a family which is differentiable in quadratic mean (van der Vaart (1998)). Moreover, none of the groups should vanish asymptotically, i.e. for $k = T, R, P$ and $n = n_T + n_R + n_P$

$$\frac{n_k}{n} \longrightarrow w_k \tag{3.1}$$

holds for $n_R, n_T, n_P \to \infty$ and some $w_k \in \,]0, 1[$, the (asymptotic) proportion of the numbers of patients in group $k = T, R, P$.

3.1 Retention of Effect Wald-type Test (RET)

In order to come up with a test for (1.3) we rewrite this as

$$H_{0,h(\theta_k)}: \quad \eta := h(\theta_T) - \Delta \, h(\theta_R) + (\Delta - 1)h(\theta_P) \quad \leq \quad 0 \,. \tag{3.2}$$

The MLE of $h(\theta_k)$, $k = T, R, P$, is obtained by plugging in the MLE $\hat{\theta}_k$ of θ_k, which is well-defined and asymptotically normally distributed by assumption. By the delta-method this

yields that $\sqrt{n_k}(h(\hat{\theta}_k) - h(\theta_k))$ is centered asymptotically normally distributed with variance

$$\sigma_k^2 = \left(\frac{\partial}{\partial\theta} h(\theta_k)\right) \cdot I(\theta_k)^{-1} \cdot \left(\frac{\partial}{\partial\theta} h(\theta_k)\right)^T$$

and I the Fisher-information-matrix, i.e.

$$I(\theta) = -E_\theta \left[\frac{\partial^2}{\partial^2\theta} \log f(\theta, X) \right] .$$

Hence, the linear contrast $\sqrt{n}(\hat{\eta} - \eta)$, where the MLE of η is obtained by plugging in the MLE's $\hat{\theta}_k$, $k = R, T, P$, in the left hand side of (3.2), is centered asymptotically normal with variance

$$\sigma^2 = \frac{\sigma_T^2}{w_T} + \frac{\Delta^2 \sigma_R^2}{w_R} + \frac{(1-\Delta)^2 \sigma_P^2}{w_P} . \tag{3.3}$$

As we have mentioned in the introduction estimation of σ^2 simply by the MLE often leads to an unsatisfactory approximation of the asymptotic normal law and various improvements have been suggested in specific settings, mainly for the case of binary endpoints (see next section). Therefore, we will treat the case of restricted maximum likelihood estimation as well. To this end, let $\hat{\sigma}_{ML}^2$ denote the MLE of σ^2 and $\hat{\sigma}_{RML}^2$ denote the MLE with restriction to the null hypothesis, i.e. the MLE of σ^2 under the restriction in (3.2). Further, let $\hat{\sigma}^2$ either denote $\hat{\sigma}_{ML}^2$ or $\hat{\sigma}_{RML}^2$, see the next Section 3.2.1 for a discussion of both estimators. Both estimators are consistent under the null hypothesis. Thus, we obtain in order to test (3.2) as a test-statistic

$$T = \frac{\sqrt{n} \cdot \hat{\eta}}{\hat{\sigma}} = \sqrt{n} \cdot \frac{h(\hat{\theta}_T) - \Delta\, h(\hat{\theta}_R) + (\Delta - 1)h(\hat{\theta}_P)}{\hat{\sigma}} \tag{3.4}$$

which is asymptotically standard normally distributed. Therefore, $H_{0,h(\theta_k)}$ is to reject if $T > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $1 - \alpha$-quantile of the standard normal distribution and α a specified significance level. Due to the formulation of the hypothesis and the test decision we will denote this test by *Retention of Effect Wald Test (RET)*. The procedure to perform the RET is summarized in Figure 3.1.

3.2 The estimators of the asymptotic variance σ^2 and their limits

In some situation, e.g. for normally distributed endpoints, it is sufficient to estimate the asymptotic variance in (3.3) by the (unrestricted) MLE (see Pigeot et al. (2003)). Roughly speaking, this is due to the fact that the asymptotic variance of the test statistic does not depend on the parameters $h(\theta_k)$ (in the normal case the mean) which only enter into the

hypothesis. However, e.g. for the case of binary endpoints the variance depends on the success probabilities itself and an improvement in the accuracy of the asymptotic normality can be obtained by estimation restrictedly to the null hypothesis. This fact has been pointed out by Farrington and Manning (1990) for the two sample comparison with binomial endpoints and various improvements have been suggested since (see e.g. Chan (1998), Röhmel and Mansmann (1999), Skipka et al. (2004)). For the retention of effect hypothesis Kieser and Friede (2007) demonstrate in an extensive simulation study that the restricted Wald-type test (Farrington and Manning (1990) adjustment) works satisfactorily and clearly outperforms the unrestricted Wald-type test concerning the accuracy of the nominal level.

3.2.1 Computation of $\hat{\sigma}^2_{ML}$ and $\hat{\sigma}^2_{RML}$

Typically, the variance σ^2 is a continuous function of the parameters θ_k, $k = T, R, P$. Thus, the MLE $\hat{\sigma}^2_{ML}$ is obtained by plugging the MLE's $\hat{\theta}_k$

$$\hat{\sigma}_{ML} = \sigma(\hat{\theta}_T, \hat{\theta}_R, \hat{\theta}_P).$$

In order to obtain the restricted MLE $\hat{\sigma}^2_{RML}$ the $\hat{\theta}_k$'s have to be replaced by their restricted versions, i.e.

$$\hat{\sigma}_{RML} = \sigma(\hat{\theta}_{T,H_0}, \hat{\theta}_{R,H_0}, \hat{\theta}_{P,H_0})$$

with

$$(\hat{\theta}_{T,H_0}, \hat{\theta}_{R,H_0}, \hat{\theta}_{P,H_0}) = \arg\sup_{(\theta_T, \theta_R, \theta_P) \in H_{0,h(\theta_k)}} \sum_{k=T,R,P} \sum_{i=1}^{n_k} \log f(\theta_k, x_{ki}). \qquad (3.5)$$

The restricted MLEs $(\hat{\theta}_{T,H_0}, \hat{\theta}_{R,H_0}, \hat{\theta}_{P,H_0})$ can be computed in the following way: if the unrestricted MLEs $\hat{\theta}_k$, $k = T, R, P$, are located in $H_{0,h(\theta_k)}$, i.e. $h(\hat{\theta}_T) - \Delta h(\hat{\theta}_R) + (\Delta - 1)h(\hat{\theta}_P) \le 0$, they coincide with the restricted MLEs. Otherwise the restricted MLEs can be determined by restricting the likelihood function to the boundary of $H_{0,h(\theta_k)}$ by means of substituting $\theta_T = h^{-1}(\Delta h(\theta_R) + (1 - \Delta)h(\theta_P))$ in the common likelihood function (left hand side from (3.5)) and maximizing this with respect to θ_R and θ_P numerically or, if possible, analytically.

3.2.2 Limits of the variance estimators

The limits of the MLEs $\hat{\sigma}^2_{ML}$ and $\hat{\sigma}^2_{RML}$ are fundamental for sample size planning in the following Section 4. For the derivation of the limits let us denote the true (unknown) parameters by $\theta_k^{(0)}$, $k = T, R, P$, and correspondingly $\eta^{(0)} = h(\theta_T^{(0)}) - \Delta h(\theta_R^{(0)}) + (\Delta - 1)h(\theta_P^{(0)})$ and

$$\sigma^2_{0,k} = \left(\frac{\partial}{\partial \theta} h(\theta_k^{(0)}) \right) \cdot I(\theta_k^{(0)})^{-1} \cdot \left(\frac{\partial}{\partial \theta} h(\theta_k^{(0)}) \right)^T \qquad (3.6)$$

for $k = R, T, P$ and

$$\sigma_0^2 \;=\; \frac{\sigma_{0,T}^2}{w_T} + \frac{\Delta^2 \sigma_{0,R}^2}{w_R} + \frac{(1-\Delta)^2 \sigma_{0,P}^2}{w_P} \;. \tag{3.7}$$

The unrestricted MLE $\hat{\sigma}_{ML}^2$ is always a consistent estimator, i.e. $\hat{\sigma}_{ML}^2 \xrightarrow{a.s.} \sigma_0^2$ as $n \to \infty$. However, the restricted MLE $\hat{\sigma}_{RML}^2$ is only consistent when the true parameters are located in the hypothesis, i.e. $\eta^{(0)} \leq 0$. In other words, the limit of $\hat{\sigma}_{RML}^2$ is no more equal to σ_0^2, in general. We will now derive the limit of the restricted MLE $\hat{\sigma}_{RML}^2$, when the parameters are located in the alternative, i.e. $\eta^{(0)} > 0$. This requires computation of the Kullback-Leibler-divergence (KL-divergence) between two parameter constellations. To this end, let $\zeta = (\theta_T, \theta_R, \theta_P)$ denote any parameter in the parameter space $\Theta^3 \subseteq \mathbb{R}^{3d}$ and $\zeta^{(0)}$ the true parameter. Then we define for the three-sample case a weighted KL-divergence between $\zeta^{(0)}$ and ζ with weights $c = (c_T, c_R, c_P)$ by

$$K(\zeta^{(0)}, \zeta, c) = \sum_{k=T,R,P} c_k \cdot K(\theta_k^{(0)}, \theta_k) \;, \tag{3.8}$$

where $K(\theta_k^{(0)}, \theta_k) = E_{\theta_k^{(0)}}[\log f(\theta_k^{(0)}, X) - \log f(\theta_k, X)]$ denotes the usual KL-divergence measuring the difference between two densities.

The next theorem shows that the restricted MLE denoted by $\hat{\zeta}_{H_0} = (\hat{\theta}_{T,H_0}, \hat{\theta}_{R,H_0}, \hat{\theta}_{P,H_0})$ converges to the minimizer of the sample size weighted Kullback-Leibler-divergence (KL-divergence) with respect to the true parameter, i.e.

$$\hat{\zeta}_{H_0} \xrightarrow{a.s.} \zeta_{H_0}$$

with

$$\zeta_{H_0} = (\theta_{T,H_0}, \theta_{R,H_0}, \theta_{P,H_0}) := \arg\min_{\zeta \in H_0} K(\zeta^{(0)}, \zeta, (w_T, w_R, w_P)).$$

For this theorem we require two further assumptions. In Assumption 1 we require the minimizer to be well-defined and Assumption 2 ensures that the likelihood is bounded, i.e. we do not end up with a unbounded likelihood as for example in the case of a mixture of two normal distributions.

Assumption 1: For $\zeta^{(0)}$ in the alternative H_1 and $n_k/n \to w_k \in]0,1[$, $w = (w_T, w_R, w_P)$, the minimum $\zeta_{H_0} = (\theta_{T,H_0}, \theta_{R,H_0}, \theta_{P,H_0}) = \arg\min_{\zeta \in H_0} K(\zeta^{(0)}, \zeta, w)$ is well-defined.

Assumption 2: For any sequence $\zeta^{(n)} = (\theta_T^{(n)}, \theta_R^{(n)}, \theta_P^{(n)})$ in H_0 with $\lim_{n \to \infty} \zeta^{(n)}$ in $\overline{\Theta}^3 \setminus \Theta^3$ or with $\lim_{n \to \infty} \| \zeta^{(n)} \| = \infty$

$$\lim_{n \to \infty} \prod_{k=T,R,P} f(\theta_k^{(n)}, x_k) \;=\; 0$$

holds $P^{\zeta^{(0)}}$ almost everywhere.

Theorem 1: *Let $\hat{\zeta}_n^{H_0}$ denote the MLE restricted to H_0. Then under the Assumptions 1 and 2*

$$\hat{\zeta}_n^{H_0} \xrightarrow{a.s.} \zeta_{H_0} .$$

Proof. Let

$$Q_n(\zeta) = - \sum_{k=T,R,P} \frac{1}{n} \sum_{i=1}^{n_k} \log f(X_{ki}, \theta_k)$$

and

$$Q(\zeta) = - \sum_{k=T,R,P} w_k \cdot E_{\theta_k^{(0)}}[\log f(X_{k1}, \theta_k)].$$

Note that by definition

$$K(\zeta^{(0)}, \zeta, w) = Q(\zeta) - Q(\zeta^{(0)})$$

holds and consequently $\zeta_{H_0} = \arg\min_{\zeta \in H_0} K(\zeta^{(0)}, \zeta, w)$ is also the well-defined minimizer of $Q(\zeta)$ in H_0.

Assumption 2 ensures that the MLE is asymptotically almost surely located in a compact set, i.e. there exists compact subset $\tilde{H}_0$ such that

$$\lim_{n\to\infty} \hat{\zeta}_n^{H_0} = \lim_{n\to\infty} \hat{\zeta}_n^{\tilde{H}_0} \qquad a.s.$$

A proof for $\lim_{n\to\infty} \| \zeta^{(n)} \| = \infty$ can be found in Wald (1949). However, for $\lim_{n\to\infty} \zeta^{(n)}$ in $\overline{\Theta}^3 \setminus \Theta^3$ this can be proved analogously. Hence, we assume w.l.o.g. that H_0 is compact. Therefore, the convergence

$$Q_n(\zeta) \xrightarrow{a.s.} Q(\zeta)$$

is uniformly in H_0 (see Jennrich, 1969, Theorem 2) and we can apply Lemma 2.2 from White (1980), which yields that $\hat{\zeta}_n^{H_0} = \arg\min_{\zeta \in H_0} Q_n(\zeta)$ converges almost surely to the well-defined minimum ζ_{H_0} of $Q(\zeta)$ in H_0. $\qquad\square$

Therefore, the limit of the restricted MLE $\hat{\sigma}_{RML}^2$ is obtained by

$$\sigma_{RML}^2 = \frac{\sigma_{T,H_0}^2}{w_T} + \frac{\Delta^2 \sigma_{R,H_0}^2}{w_R} + \frac{(1-\Delta)^2 \sigma_{P,H_0}^2}{w_P} \tag{3.9}$$

with

$$\sigma_{k,H_0}^2 = \left(\frac{\partial}{\partial\theta} h(\theta_{k,H_0}) \right) \cdot I(\theta_{k,H_0})^{-1} \cdot \left(\frac{\partial}{\partial\theta} h(\theta_{k,H_0}) \right)^T \tag{3.10}$$

for $k = T, R, P$.

Performing the RET:

Input: $h(\cdot)$ Measure of efficacy

Δ Non-inferiority margin

α Significance level

$X_{T1}, \ldots, X_{Tn_T}$ sample from the test treatment group

$X_{R1}, \ldots, X_{Rn_R}$ sample from the reference treatment group

$X_{P1}, \ldots, X_{Pn_P}$ sample from the placebo group

Procedure:

1. Compute the MLE $\hat{\theta}_k$ for each group $k = T, R, P$.

2. Determine
$$\hat{\eta} = h(\hat{\theta}_T) - \Delta\, h(\hat{\theta}_R) + (\Delta - 1)h(\hat{\theta}_P)$$

3. Compute $\hat{\sigma}$, see Section 3.2.1.

4. Compute the test statistic as
$$T = \sqrt{n_T + n_R + n_P} \cdot \frac{\hat{\eta}}{\hat{\sigma}}$$

5. Reject $H_{0,h(\theta_k)}$ if $T > z_{1-\alpha}$ (quantile of the standard normal distribution).

Figure 3.1: *Performing the RET.*

3.2.3   Numerical computation of σ_{RML}

For computing the minimizers θ_{k,H_0}, $k = T, R, P$, and therewith σ_{RML} for a parameter constellation in the alternative, i.e. $\eta^{(0)} > 0$, it is sufficient to restrict to the boundary of $H_{0,h(\theta_k)}$, i.e. we replace in the weighted KL-divergence (3.8) θ_T by $h^{-1}(\Delta h(\theta_R) + (1-\Delta)h(\theta_P))$ and then minimize the KL-divergence with respect to θ_R and θ_P.

In practice, the analytic solution to the minimization problem of the KL-divergence may be hard (confer the example of Poisson endpoints in Section 5.2.3) or even unfeasible to find. In this case, numerical minimization becomes necessary. To this end, it is important to note that the minimization of the KL-divergence often results in a convex optimization problem and fast algorithms for convex optimization, such as the Newton-Raphson algorithm, become feasible. The following theorem states conditions to obtain a convex optimization problem.

Theorem 2: *Let* $-E_{\theta_k^{(0)}}[\frac{\partial^2}{\partial^2\theta} \log f(\theta, X)]$ *be non-negative for all* $\theta \in \Theta$ *and* $\theta_k^{(0)}$, $k = T, R, P$ *and let* Θ *be a convex set. Further, let* $h^{-1}(\Delta h(\theta_R) + (1-\Delta)h(\theta_P))$ *be an affine transformation in* θ_R *and* θ_P. *Then the minimization in* ζ *of the weighted KL-divergence (3.8) with restriction to the boundary of the null hypothesis is a convex optimization problem.*

Proof. The condition $-E_{\theta_k^{(0)}}[\frac{\partial^2}{\partial^2\theta} \log f(\theta, X)]$ ensures that the KL-divergence $K(\theta_k^{(0)}, \theta)$ is a

convex function in θ for $\theta_k^{(0)}$, $k = T, R, P$. Thus, the weighted KL-divergence

$$K(\zeta^{(0)}, (\theta_T, \theta_R, \theta_P), c) = \sum_{k=T,R,P} c_k \cdot K(\theta_k^{(0)}, \theta_k)$$

is convex in the arguments θ_k, $k = T, R, P$. Let us denote

$$g(\theta_R, \theta_P) = h^{-1}(\Delta h(\theta_R) + (1 - \Delta)h(\theta_P))$$

which is an affine transformation in both arguments by assumption. Hence,

$$K(\theta_T^{(0)}, g(\theta_R, \theta_P))$$

is a convex function in (θ_R, θ_P). Therefore, the weighted KL-divergence with restriction to the boundary of the null hypothesis $H_{0,h(\theta_k)}$ represented by

$$K(\zeta^{(0)}, (g(\theta_R, \theta_P), \theta_R, \theta_P), c) = c_T \cdot K(\theta_T^{(0)}, g(\theta_R, \theta_P)) + c_R \cdot K(\theta_R^{(0)}, \theta_R) + c_P \cdot K(\theta_P^{(0)}, \theta_P)$$

is a linear combination of convex function and therewith convex in (θ_R, θ_P), again.

$$\square$$

The conditions of Theorem 1 are fulfilled in our examples of Poisson and binary endpoints which will be revisited in Section 5.

3.3 Approximating the power function of the RET

The asymptotic normality used in Section 3.1 to derive the RET is valid for parameter constellations in the hypothesis as well as for constellations in the alternative. Thus, if the variance σ^2 is estimated unrestrictedly, $\hat{\sigma}^2 = \hat{\sigma}_{ML}^2$, we obtain as an approximation to the power function of the RET, i.e. the probability of rejecting the hypothesis $H_{0,h(\theta_k)}$ in (1.3), by

$$P_{\eta^{(0)}}(T > z_{1-\alpha}) \approx 1 - \Phi\left(z_{1-\alpha} - \sqrt{n}\,\frac{\eta^{(0)}}{\sigma_0}\right). \tag{3.11}$$

However, estimating the variance σ^2 restricted to the null hypothesis, i.e. $\hat{\sigma}^2 = \hat{\sigma}_{RML}^2$, complicates the issue and changes the power function to

$$\begin{aligned} P_{\eta^{(0)}}(T > z_{1-\alpha}) &= P_{\eta^{(0)}}\left(T \cdot \frac{\sigma_{RML}}{\sigma_0} - \sqrt{n}\,\frac{\eta^{(0)}}{\sigma_0} > z_{1-\alpha} \cdot \frac{\sigma_{RML}}{\sigma_0} - \sqrt{n}\,\frac{\eta^{(0)}}{\sigma_0}\right) \\ &\approx 1 - \Phi\left(z_{1-\alpha} \cdot \frac{\sigma_{RML}}{\sigma_0} - \sqrt{n}\,\frac{\eta^{(0)}}{\sigma_0}\right). \end{aligned} \tag{3.12}$$

Note that (3.11) can be obtained from (3.12) by means of substituting σ_{RML} by σ_0.

THEORY: OPTIMAL SAMPLE ALLOCATION AND GENERAL STRATEGY FOR SAMPLE SIZE PLANNING (GSSP)

In this section we present a sample size formula for the test of the generalized retention of effect hypothesis $H_{0,h(\theta_k)}$ (1.3) introduced in Section 1. In particular, we derive the optimal allocation of the samples to the groups T, R and P in terms of maximizing the power of the RET under any fixed alternative $\eta^{(0)}$.

4.1 Optimal sample allocation

In planning a trial, one typically specifies a parameter constellation $\eta^{(0)}$ in the alternative. Our aim in this section is to optimize the allocation of samples, represented through w_k, $k = T, R, P$, as in (3.1), such that the power of the test decision in (3.11) or (3.12), respectively, is maximized. The power depends on the allocation through σ_0^2 and σ_{RML}^2.

4.1.1 Unrestricted estimation of the variance σ^2

When the variance σ^2 is estimated unrestricted in the test procedure ($\hat{\sigma}^2 = \hat{\sigma}_{ML}^2$) we have only to consider σ_0^2 to investigate the influence of the allocation on the power, confer (3.11). This means that we have to minimize σ_0^2 in order to maximize the power.

To this end, we substitute $w_P = 1 - w_T - w_R$ in (3.7), because the sample allocation has to fulfill $w_T + w_R + w_P = 1$, and obtain

$$\sigma_0^2 = \frac{\sigma_{0,T}^2}{w_T} + \frac{\Delta^2 \cdot \sigma_{0,R}^2}{w_R} + \frac{(1 - \Delta)^2 \cdot \sigma_{0,P}^2}{1 - w_T - w_R} \, .$$

Note that σ_0^2 is a convex function in (w_T, w_R). Evaluating the derivatives of σ_0^2 w.r.t. w_T and

w_R at zero yields

$$\frac{\partial}{\partial w_T} \sigma_0^2 = \frac{(1-\Delta)^2 \cdot \sigma_{0,P}^2}{(1-w_R-w_T)^2} - \frac{\sigma_{0,T}^2}{w_T^2} = 0$$

$$\frac{\partial}{\partial w_R} \sigma_0^2 = \frac{(1-\Delta)^2 \cdot \sigma_{0,P}^2}{(1-w_R-w_T)^2} - \frac{\Delta^2 \cdot \sigma_{0,T}^2}{w_T^2} = 0.$$

Solving the equations for w_T and w_R yields the minimizer

$$w_T^* = \frac{\sigma_{0,T}}{\sigma_{0,T} + \Delta \cdot \sigma_{0,R} + |1-\Delta| \cdot \sigma_{0,P}}$$

$$w_R^* = \frac{\Delta \cdot \sigma_{0,R}}{\sigma_{0,T} + \Delta \cdot \sigma_{0,R} + |1-\Delta| \cdot \sigma_{0,P}}$$

and therewith

$$w_P^* = \frac{|1-\Delta| \cdot \sigma_{0,P}}{\sigma_{0,T} + \Delta \cdot \sigma_{0,R} + |1-\Delta| \cdot \sigma_{0,P}} .$$

Thus, the optimal allocation in terms of minimizing the variance σ_0^2 is given by

$$n_T^* : n_R^* : n_P^* = 1 : \Delta \frac{\sigma_{0,R}}{\sigma_{0,T}} : |1-\Delta| \frac{\sigma_{0,P}}{\sigma_{0,T}}. \tag{4.1}$$

The resulting optimal minimal variance is given by

$$\sigma_{0,optimal}^2 = \left(\sigma_{0,T} + \Delta\sigma_{0,R} + |1-\Delta|\sigma_{0,P}\right)^2 .$$

Remark: For the specific case of normal endpoints with equal variances (Pigeot et al., 2003; Schwartz and Denne, 2006) we obtain the optimal allocation $1 : \Delta : |1-\Delta|$, again.

4.1.2 Restricted estimation of the variance σ^2

When the variance σ^2 is estimated under restriction to $H_{0,h(\theta_k)}$ the asymptotic power in (3.12) depends additionally on σ_{RML}/σ_0 because under any alternative the restricted estimator $\hat{\sigma}_{RML}$ is not a consistent estimator for σ_0. Nevertheless, the asymptotically optimal allocation derived for the unrestricted case in Section 4.1.1 is again optimal in an asymptotic sense because the power in (3.12) is dominated by the term $\sqrt{n} \cdot \eta(0)/\sigma_0$ as n grows. Hence, the allocation (4.1) derived in the previous section, which minimizes the variance σ_0, is also the (asymptotically) optimal allocation in terms of maximizing the power when the variance σ^2 is estimated restricted to $H_{0,h(\theta_k)}$.

Remark: (a) We would like to stress that this result can be applied to the case of binary endpoints (see Section 5.1.2). This leads to different results as in Kieser and Friede (2007) who derived the optimal allocation under the additional restriction of a fixed ratio w_R/w_T.

Optimal allocation:

Input: $h(\cdot)$ Measure of efficacy

$\theta_T^{(0)}, \theta_R^{(0)}, \theta_P^{(0)}$ Parameter constellation in the alternative, $\eta^{(0)} > 0$.

Δ Non-inferiority margin

Procedure:

1. Compute $\sigma_{0,k}^2$, $k = T, R, P$, via (3.6)

$$\sigma_{0,k}^2 = \left(\frac{\partial}{\partial \theta} h(\theta_k^{(0)}) \right) \cdot I(\theta_k^{(0)})^{-1} \cdot \left(\frac{\partial}{\partial \theta} h(\theta_k^{(0)}) \right)^T$$

2. Determine the optimal sample allocation via (4.1)

$$n_T^* : n_R^* : n_P^* \; = \; 1 : \Delta \, \frac{\sigma_{0,R}}{\sigma_{0,T}} : |1 - \Delta| \, \frac{\sigma_{0,P}}{\sigma_{0,T}}$$
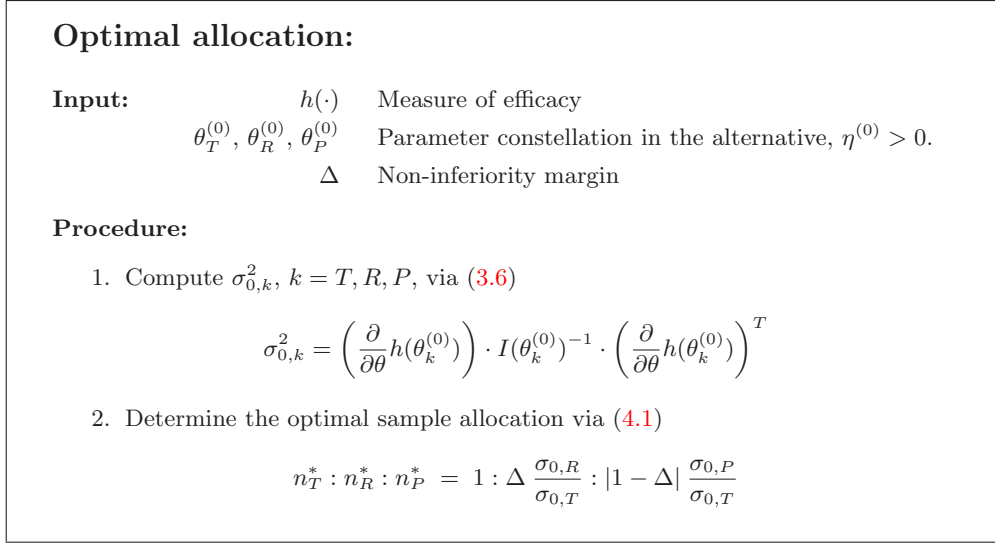
Figure 4.1: *Summary of the procedure to determine the optimal sample allocation.*

(b) The asymptotically optimal allocation presented in (4.1) should be understood as approximative for finite samples as it is customary for asymptotic results. Nevertheless, for in this paper presented examples we will show in Section 5 that the optimal allocation is also very accurate for finite samples, e.g. for a power of 80%. However, one should be aware of the fact that, it is not guaranteed that the allocation (4.1) is optimal, in general, in particular for small sample sizes.

4.1.3   Rule of thumb

The asymptotically optimal sample allocation (4.1) depends on the choice of the alternative $\zeta^{(0)} > 0$. If one is not clear about the choice of the alternative or wants to consider more than one alternative, we recommend to use as a rule of thumb the allocation $1 : \Delta : (1 - \Delta)$. We show for $\theta_R^{(0)} = \theta_T^{(0)}$ and $0 \leq \Delta \leq 1$ in Theorem 3 and 4 that the allocation $1 : \Delta : (1 - \Delta)$ is more appropriate than the commonly used 2:2:1 allocation (the balanced allocation) if $\sigma_{0,P}^2 / \sigma_{0,T}^2 < 1 + \sqrt{5/4} \, (1 + \sqrt{3})$. Note that a lower bound for $\sigma_{0,P}^2 / \sigma_{0,T}^2$ is not required. Moreover, this result is valid independent of the distribution of the endpoints and of the formulation of the retention of effect hypothesis.

Theorem 3:   *If $\theta_R^{(0)} = \theta_T^{(0)}$ and $\sigma_{0,P}^2 / \sigma_{0,T}^2 < 1 + \sqrt{5/4} \approx 2.12$ then the allocation $1 : \Delta : (1 - \Delta)$ results in a smaller variance σ_0^2 (3.7) (and hence larger asymptotic power) than the allocation 2:2:1 for any $0 \leq \Delta \leq 1$.*

Proof. Substituting the allocation 2:2:1 and $1 : \Delta : (1 - \Delta)$, respectively, and $\theta_R^{(0)} = \theta_T^{(0)}$ in

the variance σ_0^2 from (3.7) yields

$$\sigma_{2:2:1}^2 = \frac{5 + 5\Delta^2}{2}\,\sigma_{0,T}^2 \;+\; 5(1-\Delta)^2\,\sigma_{0,P}$$

and

$$\sigma_{1:\Delta:1-\Delta}^2 = 2(1+\Delta)\,\sigma_{0,T}^2 \;+\; 2(1-\Delta)\,\sigma_{0,P}^2.$$

Thus, we obtain with $r := \sigma_{0,P}^2/\sigma_{0,T}^2 > 0$

$$g(\Delta, r) := \frac{\sigma_{2:2:1}^2 - \sigma_{1:\Delta:1-\Delta}^2}{\sigma_{0,T}^2} = (2.5 + 5 \cdot r)\,\Delta^2 + (-2 - 8 \cdot r)\,\Delta + (0.5 + 3 \cdot r)\,,$$

which is as a quadratic function in Δ with minimum

$$(a(r), b(r)) = \left(\frac{2 + 8 \cdot r}{5 + 10 \cdot r}\,,\; \frac{-4 \cdot (r - 1 - \sqrt{5/4})(r - 1 + \sqrt{5/4})}{10 + 20 \cdot r} \right),$$

where $0 < a(r) < 1$ and $b(r) > 0$ for $r < 1 + \sqrt{5/4}$. Thus, we obtain for $r = \sigma_{0,P}^2/\sigma_{0,T}^2 < 1 + \sqrt{5/4}$ that $g(\Delta, r) > 0$, which implies $\sigma_{2:2:1}^2 > \sigma_{1:\Delta:1-\Delta}^2$, for any $0 \leq \Delta \leq 1$.

$\square$

Theorem 4: *If $\theta_R^{(0)} = \theta_T^{(0)}$ and $\sigma_{0,P}^2/\sigma_{0,T}^2 < 1 + \sqrt{3} \approx 2.73$ then the allocation $1 : \Delta : (1-\Delta)$ results in a smaller variance σ_0^2 (3.7) (and hence larger asymptotic power) than the balanced allocation 1:1:1 for any $0 \leq \Delta \leq 1$.*

Proof. Substituting the allocation 1:1:1 and $1 : \Delta : (1 - \Delta)$, respectively, and $\theta_R^{(0)} = \theta_T^{(0)}$ in the variance σ_0^2 from (3.7) yields

$$\sigma_{1:1:1}^2 = (3 + 3\Delta^2)\,\sigma_{0,T}^2 \;+\; 3(1-\Delta)^2\,\sigma_{0,P}$$

and

$$\sigma_{1:\Delta:1-\Delta}^2 = 2(1+\Delta)\,\sigma_{0,T}^2 \;+\; 2(1-\Delta)\,\sigma_{0,P}^2.$$

Thus, we obtain with $r := \sigma_{0,P}^2/\sigma_{0,T}^2 > 0$

$$g(\Delta, r) := \frac{\sigma_{1:1:1}^2 - \sigma_{1:\Delta:1-\Delta}^2}{\sigma_{0,T}^2} = (3 + 3 \cdot r)\,\Delta^2 + (-2 - 4 \cdot r)\,\Delta + (1 + r)\,,$$

which is as a quadratic function in Δ with minimum

$$(a(r), b(r)) = \left(\frac{2 + 4 \cdot r}{6 + 6 \cdot r}\,,\; \frac{-(r - 1 + \sqrt{3})(r - 1 - \sqrt{3})}{3 + 3 \cdot r} \right),$$

where $0 < a(r) < 1$ and $b(r) > 0$ for $r < 1 + \sqrt{3}$. Thus, we obtain for $r = \sigma_{0,P}^2/\sigma_{0,T}^2 < 1 + \sqrt{3}$ that $g(\Delta, r) > 0$, which implies $\sigma_{1:1:1}^2 > \sigma_{1:\Delta:1-\Delta}^2$, for any $0 \leq \Delta \leq 1$.

$\square$

4.2 Sample size computation

When the variance σ^2 is estimated unrestrictedly ($\hat{\sigma}^2 = \hat{\sigma}_{ML}^2$) we end up with the simplified power formula (3.11). Thus, the minimal required total sample size to obtain a power of $1 - \beta$ for a given alternative $\eta^{(0)} > 0$ is determined by

$$n_{1-\beta} \approx (z_{1-\alpha} + z_{1-\beta})^2 \cdot \left(\frac{\sigma_0}{\eta^{(0)}}\right)^2 \tag{4.2}$$

with σ_0 defined in (3.7). When the variance σ^2 is estimated restricted to the null hypothesis ($\hat{\sigma}^2 = \hat{\sigma}_{RML}^2$) the sample size formula has to be derived from (3.12) and becomes more involved, viz.

$$n_{1-\beta} \approx \left(\frac{z_{1-\alpha} \cdot \sigma_{RML} + z_{1-\beta} \cdot \sigma_0}{\eta^{(0)}}\right)^2 = \left(z_{1-\alpha} \cdot \frac{\sigma_{RML}}{\sigma_0} + z_{1-\beta}\right)^2 \cdot \left(\frac{\sigma_0}{\eta^{(0)}}\right)^2 \tag{4.3}$$

with σ_{RML} derived in (3.9). As we will see the additional term σ_{RML}/σ_0 has a relevant impact on the sample size planning.

In Figure 4.2 we have summarized the general strategy for sample size planning (GSSP) for the RET when the variance σ^2 is estimated with restriction to the null hypothesis. When the variance σ^2 is estimated unrestrictedly by $\hat{\theta}_{ML}$ we may omit the steps 2.-4. in Figure 4.2 and use the simpler formula (4.2) in step 5. to compute the required sample size $n_{1-\beta}$.

Remark: We again stress that the use of σ_{RML} will significantly affect the planning of the trial. If one replaces in (4.3) σ_{RML} by σ_0 this may result in a too small or too large required sample size depending on the ratio σ_{RML}/σ_0. If the ratio σ_{RML}/σ_0 is greater (smaller) than one, then we end up with a too small (large) required sample size, i.e. the resulting power is smaller (larger) than the desired power $1 - \beta$. For example, this will be the case for Poisson distributed endpoints (see Section 2.2) and the hypothesis (2.4). We will see in Section 5.2 that $\sigma_{RML}/\sigma_0 > 1$ for all parameter constellations. In contrast, for binary endpoints (see Section 2.1) and the hypothesis (2.1), we will show in Section 5.1 that there is no strict relationship between σ_{RML} and σ_0. Thus, a wrongly specified sample size may result in a too large or too small power compared to the aspired one.

General strategy for sample size planning (GSSP)

Input:

	$h(\cdot)$	Measure of efficacy
	$\theta_T^{(0)}, \theta_R^{(0)}, \theta_P^{(0)}$	Parameter constellation in the alternative, $\eta^{(0)} > 0$.
	$w_T : w_R : w_P$	Allocation of samples
	Δ	Non-inferiority margin
	α	Significance level
	$1 - \beta$	Aspired power

Procedure:

1. Compute $\eta^{(0)} = h(\theta_T^{(0)}) - \Delta\, h(\theta_R^{(0)}) + (\Delta - 1)h(\theta_P^{(0)})$.

2. Compute σ_0^2 via (3.7)

$$\sigma_0^2 = \frac{\sigma_{0,T}^2}{w_T} + \frac{\Delta^2 \cdot \sigma_{0,R}^2}{w_R} + \frac{(1 - \Delta)^2 \cdot \sigma_{0,P}^2}{w_P}$$

   with $\sigma_{0,k}^2$, $k = T, R, P$, from (3.6)

$$\sigma_{0,k}^2 = \left(\frac{\partial}{\partial \theta} h(\theta_k^{(0)}) \right) \cdot I(\theta_k^{(0)})^{-1} \cdot \left(\frac{\partial}{\partial \theta} h(\theta_k^{(0)}) \right)^T.$$

3. Determine the weighted KL-divergence (3.8) for the endpoint of investigation.

4. Compute the parameter constellation θ_{k,H_0}, $k = T, R, P$, in the null hypothesis, which minimizes the weighted KL-divergence to the true parameter. This can be done analytically or numerically (confer Section 3.2.3).

5. Compute

$$\sigma_{RML}^2 \;=\; \frac{\sigma_{T,H_0}^2}{w_T} + \frac{\Delta^2 \sigma_{R,H_0}^2}{w_R} + \frac{(1 - \Delta)^2 \sigma_{P,H_0}^2}{w_P}$$

   for θ_{k,H_0}, $k = T, R, P$, via (3.9) and (3.10).

6. Use formula (4.3) to compute the minimal total required sample size

$$n_{1-\beta} \;\approx\; \left(\frac{z_{1-\alpha} \cdot \sigma_{RML} + z_{1-\beta} \cdot \sigma_0}{\eta^{(0)}} \right)^2.$$

Figure 4.2: *General strategy for sample size planning (GSSP) when the variance σ^2 is estimated with restriction to the null hypothesis.*

SPECIFIC MODELS - APPLICATIONS OF THEORY AND ILLUSTRATION OF IMPROVEMENTS

In the following we will perform the RET for the examples from Section 2 and we will illustrate the general strategy for sample size planning (GSSP) including a detailed investigation of the optimal allocation. In addition, we discuss censored, exponentially distributed endpoints by means of summarizing the results from Mielke et al. (2008) in Section 5.3.

5.1 Binary endpoints: Treatment of depression

In this section we revisit the example in the treatment of depression introduced in Section 2.1, whereat we focus on the retention of effect hypothesis with $h(\pi_k) = \pi_k$ (2.1) here to illustrate the improvements in accuracy of the procedures. The log odds retention of effect hypothesis (2.2) will be only discussed briefly in Section (5.1.4) as the procedures are quite analogue. Nevertheless, the deviation of the optimal sample allocation requires further investigations as it turns out that the finite sample approximation of the asymptotically optimal allocation does not performs well in general for the log odds retention of effect hypothesis. This will be discussed in detail in the next Section 6.

5.1.1 Performing the RET

For the sake of completeness we recall the RET for the situation $h(\pi_k) = \pi_k$, which was already introduced by Kieser and Friede (2007). The MLE of π_k is $\hat{\pi}_k = n_k^{-1} \sum_{i=1}^{n_k} X_{ki}$ which is asymptotically normally distributed with variance $\sigma_k^2 = \pi_k(1 - \pi_k)$. Hence, the unrestricted MLE of the variance σ^2 is given by (cf. (3.3)),

$$\hat{\sigma}_{ML}^2 = n \cdot \left(\frac{\hat{\pi}_T(1 - \hat{\pi}_T)}{n_T} + \Delta^2 \frac{\hat{\pi}_R(1 - \hat{\pi}_R)}{n_R} + (1 - \Delta)^2 \frac{\hat{\pi}_P(1 - \hat{\pi}_P)}{n_P} \right)$$
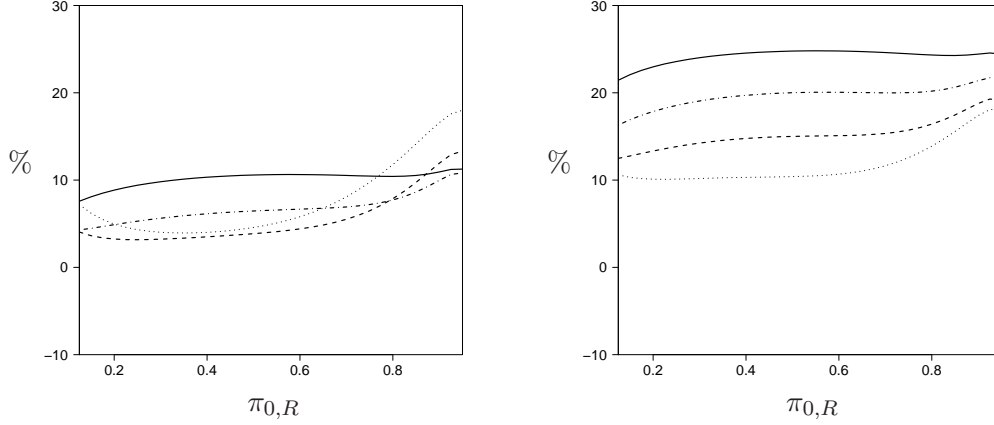
Figure 5.1: *Example of binary distributed endpoints: **Sample size reduction in %** when optimal allocation (5.3) is used instead of the balanced allocation (right figure) and instead of the allocation 2:2:1 (left figure) for $\pi_{0,P} = 0.1$ and different values $\Delta = 0.5$ (dotted line), $\Delta = 0.6$ (dashed line), $\Delta = 0.7$ (dotdash line), $\Delta = 0.8$ (solid line).*

and we end up with the test statistic (see (3.4))

$$T = \sqrt{n} \cdot \frac{\hat{\pi}_T - \Delta\,\hat{\pi}_R + (\Delta - 1)\hat{\pi}_P}{\hat{\sigma}_{ML}} \tag{5.1}$$

in order to test H_{0,π_k} in (2.1) which is rejected if $T > z_{1-\alpha}$.

Let us now consider the case where σ^2 is estimated restrictedly (cf. Farrington and Manning, 1990). The restricted version of the Wald-type test is observed by replacing the MLEs $\hat{\pi}_k$ in the denominator by the to H_{0,π_k} restricted ones. Here, we have computed the restricted MLEs accordingly to Section 3.2.1: if the unrestricted MLEs $\hat{\pi}_k$, $k = T, R, P$, are located in H_{0,π_k}, i.e. $\hat{\pi}_T - \Delta\,\hat{\pi}_R + (\Delta - 1)\hat{\pi}_P < 0$, they coincide with the restricted MLEs. Otherwise we have determined the restricted MLEs by restricting the likelihood function to the boundary of H_{0,π_k} by means of substituting $\pi_T = \Delta\pi_R + (1 - \Delta)\pi_P$ in the common likelihood function and maximizing this with respect to π_R and π_P numerically. Note, that in contrast to the two-sample case (Farrington and Manning, 1990), an analytical computation of the restricted MLE's is not feasible, anymore.

The RET for the hypothesis (2.1) with $\Delta = 0.8$ yields $T = 2.104$ (2.108) in (5.1) using the restricted (unrestricted) estimator for the variance estimation and corresponding p-values 1.77% (1.75%). Thus, we would reject H_{0,θ_k} from (2.1) in both cases and claim that the test treatment duloxetine is non-inferior over paroxetine.

Table 5.1: **Example of binomial distributed endpoints:** *Optimal sample allocation, limit of variance estimator $\hat{\sigma}_{RML}$ and required samples size from formula (4.3) and (4.2), respectively, to obtain a power of 0.7 and 0.8, respectively, when the variance σ^2 is estimated restrictedly to the null-hypothesis (unrestrictedly), where $\alpha = 5\%$, $\Delta = 0.7$.*

$\pi_{0,P}$	$\pi_{0,T}$	w_T^*	w_R^*	w_P^*	Optimal allocation			2:2:1 allocation		
					$\frac{\sigma_{RML}}{\sigma_0}$	$n_{0.7}$	$n_{0.8}$	$\frac{\sigma_{RML}}{\sigma_0}$	$n_{0.7}$	$n_{0.8}$
0.1	0.3	0.527	0.369	0.104	0.994	997 (988)	1308 (1297)	1.014	1054 (1076)	1388 (1414)
	0.5	0.532	0.372	0.096	0.986	296 (289)	387 (380)	1.006	315 (318)	415 (418)
	0.7	0.527	0.369	0.104	0.955	118 (110)	154 (145)	0.965	127 (120)	165 (158)
	0.9	0.500	0.350	0.150	0.791	43 (30)	54 (39)	0.759	48 (31)	60 (41)
0.2	0.5	0.515	0.361	0.124	0.995	552 (547)	724 (719)	1.002	582 (583)	765 (766)
	0.7	0.510	0.357	0.133	0.977	176 (170)	230 (223)	0.969	188 (179)	245 (235)
	0.9	0.476	0.333	0.190	0.835	57 (43)	72 (56)	0.795	63 (44)	79 (58)
0.3	0.5	0.506	0.354	0.139	0.998	1279 (1275)	1680 (1675)	1.012	1341 (1341)	1761 (1762)
	0.7	0.500	0.350	0.150	0.986	281 (275)	368 (361)	0.975	298 (287)	390 (377)
	0.9	0.463	0.324	0.212	0.867	76 (61)	98 (81)	0.830	84 (63)	106 (83)
0.4	0.7	0.495	0.346	0.159	0.993	504 (499)	661 (655)	0.961	532 (518)	697 (680)
	0.9	0.457	0.320	0.224	0.896	107 (91)	138 (119)	0.863	117 (93)	149 (122)
0.5	0.7	0.493	0.345	0.161	0.997	1134 (1129)	1489 (1483)	0.988	1191 (1170)	1561 (1537)
	0.9	0.455	0.318	0.227	0.924	161 (143)	209 (188)	0.894	174 (147)	224 (193)
0.6	0.7	0.495	0.346	0.159	0.999	4490 (4484)	5898 (5891)	0.994	4695 (4655)	6162 (6116)
	0.8	0.484	0.339	0.178	0.993	904 (894)	1186 (1175)	0.977	953 (921)	1247 (1209)
	0.9	0.457	0.320	0.224	0.950	272 (251)	353 (330)	0.923	292 (258)	377 (339)
0.7	0.8	0.489	0.343	0.168	0.998	3505 (3495)	4603 (4591)	0.989	3672 (3611)	4814 (4744)
	0.9	0.463	0.324	0.212	0.974	571 (549)	746 (721)	0.949	609 (562)	792 (739)
0.8	0.9	0.476	0.333	0.190	0.992	2101 (2076)	2756 (2727)	0.975	2214 (2130)	2895 (2798)

5.1.2 Optimal allocation

For binary distributed endpoints and the hypothesis (2.1) the optimal allocation of samples is given by

$$n_T^* : n_R^* : n_P^* \;=\; 1 : \Delta\sqrt{\frac{\pi_{0,R}(1-\pi_{0,R})}{\pi_{0,T}(1-\pi_{0,T})}} : |1-\Delta|\sqrt{\frac{\pi_{0,P}(1-\pi_{0,P})}{\pi_{0,T}(1-\pi_{0,T})}} \qquad (5.2)$$

according to (4.1). For the commonly used alternative $\pi_{0,R} = \pi_{0,T}$ the allocation simplifies to

$$n_T^* : n_R^* : n_P^* \;=\; 1 : \Delta : |1-\Delta|\sqrt{\frac{\pi_{0,P}(1-\pi_{0,P})}{\pi_{0,T}(1-\pi_{0,T})}} \,. \qquad (5.3)$$

In contrast to the case of normally distributed endpoints, where the optimal allocation is given by $1 : \Delta : |1-\Delta|$ (cf. Pigeot et al., 2003), the optimal allocation depends on the parameter of investigation.

Kieser and Friede (2007) derived the optimal allocation under the additional constraint that the test and reference group are balanced, $n_T^* = n_R^*$. Our result (5.2) shows that this restriction does not lead to an approximative optimal allocation, in general. Exemplary, Kieser and Friede (2007) derive that the allocation $2.1 : 2.1 : 1$ would be optimal for $\pi_P = 0.1$, $\pi_T = \pi_R = 0.9$ and $\Delta = 0.6$, whereas (5.3) yields an optimal allocation of $2.5 : 1.5 : 1$, giving more weight to the test group relative to the reference group and nearly the same to the placebo group. The allocation $2.5 : 1.5 : 1$ and the allocation $2.1 : 2.1 : 1$ result in a total required sample size of 79 and 89, respectively, when a power $1 - \beta$ of 80% is desired. Thus, our optimal allocation makes a further reduction of total sample size of about 12% possible in this specific setting. The sample size reductions which are possible in other settings are illustrated in Figure 5.1, where the reduction for the optimal allocation instead of a balanced and a 2:2:1 allocation, respectively, is presented for $\pi_{0,P} = 0.1$ and different values of Δ, exemplary. For the 2:2:1 allocation we observe reductions between about 3% and 10%. For the balanced allocation there are reductions up to 20% and more possible. Thus, the 2:2:1 allocation is more apporiate than the balanced allocation. However, it can be further improved by the optimal one (5.3).

5.1.3   Planning a trial - applying the GSSP

For binary distributed endpoints the weighted KL-divergence is given by

$$K(\zeta^{(0)}, \zeta, w) = \sum_{k=T,R,P} w_k \cdot \left(\pi_k^{(0)} \cdot \log \frac{\pi_k^{(0)}}{\pi_k} + (1 - \pi_k^{(0)}) \cdot \log \frac{1 - \pi_k^{(0)}}{1 - \pi_k} \right) \qquad (5.4)$$

with $\zeta = (\pi_T, \pi_R, \pi_P)$ and $\zeta^{(0)} = (\pi_T^{(0)}, \pi_R^{(0)}, \pi_P^{(0)})$. We restrict our investigations in the following to the commonly used alternative $\pi_T^{(0)} = \pi_R^{(0)}$. To restrict the minimization problem of the weighted KL-divergence to H_{0,π_k} (2.1) we substitute $\pi_T = \Delta \pi_R + (1 - \Delta)\pi_P$ in (5.4). We have minimized the KL-divergence (5.4) in π_R and π_P by the Newton-Raphson algorithm, confer Section 3.2.3. Note that this is a strictly convex optimization problem by Theorem 1 because

$$-E_{\pi_k^{(0)}} \left[\frac{\partial^2}{\partial^2 \pi} \log f(\pi, X) \right] = \frac{\pi_k^{(0)}}{\pi^2} + \frac{1 - \pi_k^{(0)}}{(1 - \pi)^2} > 0$$

for any π and $\pi_k^{(0)}$. This guarantees the existence of a unique minimizer and geometric convergence of the Newton-Raphson algorithm. Based on the obtained results the limit σ_{RML}^2 of the restricted MLE's of the variance is computed and compared to the true variance σ_0^2, see Table 5.1, columns 6 and 9. We used throughout Table 5.1 a choice of $\Delta = 0.7$, exemplary. We may use (4.2) and (4.3), respectively, to compute the total required samples size. The results are also displayed in Table 5.1 for a power $1 - \beta$ of 0.7 and 0.8, respectively, for the

optimal allocation, displayed in the columns 3-5 of Table 5.1, and the commonly used 2:2:1 allocation for the purpose of illustrating the influence of allocation on the total required sample size. The sample size values in brackets are determined by (4.2), i.e. the RET is performed with unrestricted estimation of variance $\hat{\sigma}^2 = \hat{\sigma}^2_{ML}$, and the values in front without brackets are determined by (4.3), i.e. the RET is performed with restricted estimation of variance $\hat{\sigma}^2 = \hat{\sigma}^2_{RML}$. For large sample sizes the differences between both values are relatively small, whereas for small to moderate sample sizes ($n < 200$) the differences are more pronounced. The amount of difference is driven by the difference between σ_{RML} and σ_0, see again Table 5.1, column 6 and 9.

It is important to note that these results differ from those obtained by Kieser and Friede (2007). This is due to the fact, that for the computation of σ^2_{RML} we have used the limit of the restricted MLE $\hat{\sigma}^2_{RML}$ instead of only choosing σ^2_{RML} such that the parameters are on the boundary of H_{0,π_k}. We will see that the usage of the exact limit σ^2_{RML} significantly improves the precision of the sample size formula (4.3). To this end, we have determined the required total sample size n via (4.3) with the usage of the exact limit σ^2_{RML} to obtain a power of 80% at level $\alpha = 2.5\%$ (in order to be comparable with the results obtained by Kieser and Friede (2007)) for different parameter settings and allocations and thereafter we have computed the resulting exact power (see Table 5.2). Note that we always have rounded down the group sample sizes n_k, $k = T, R, P$. The results obtained by Kieser and Friede (2007), who have not used the exact limit σ^2_{RML}, are presented for comparison. Kieser and Friede (2007) obtain an exact power that increases to 85% or even to 87% for some settings although $n > 200$. Whereas the power decreases up to 78% for other settings. In contrast, our method results in power values between 80% and 82% for all settings (with one exception for the case $w_T : w_R : w_P = 3 : 2 : 1$, $\Delta = 0.6$, $\pi_{0,P} = 0.1$ and $\pi_{0,R} = 0.9$ due to the small total sample size of 45). In summary, we find that our approximative formula yields very satisfactory results over a broad range of scenarios.

5.1.4 Log odds retention of effect hypothesis

For the log odds retention of effect hypothesis (2.2) the procedures modify in the following. We end up with the test statistic

$$T = \sqrt{n}\,\hat{\sigma}^{-1}\,\hat{\eta} = \sqrt{n}\,\hat{\sigma}^{-1}\left(\log\left(\frac{\hat{\pi}_T}{1-\hat{\pi}_T}\right) - \Delta\log\left(\frac{\hat{\pi}_R}{1-\hat{\pi}_R}\right) + (\Delta-1)\log\left(\frac{\hat{\pi}_P}{1-\hat{\pi}_P}\right)\right)$$

with $\hat{\pi}_k = n_k^{-1}\sum_{i=1}^{n_k} X_{ki}$ and $\hat{\sigma}$ either denoting the restricted $\hat{\sigma}_{RML}$ or the unrestricted ML-estimator $\hat{\sigma}_{ML}$ of σ given by

$$\hat{\sigma}^2_{ML} = n \cdot \left(\frac{1}{n_T\hat{\pi}_T(1-\hat{\pi}_T)} + \frac{\Delta^2}{n_R\hat{\pi}_R(1-\hat{\pi}_R)} + \frac{(1-\Delta)^2}{n_P\hat{\pi}_P(1-\hat{\pi}_P)}\right).$$

Again, $\hat{\sigma}^2_{RML}$ is observed by replacing the ML-estimators $\hat{\pi}_k$ in the denominator by the to H_0 restricted ones $\hat{\pi}_{k,H_0}$, which has to be computed numerically as already discussed for the "classical" hypothesis (2.1) for binary endpoints. Considering the data set in the treatment of depression from Section 2.1 the RET for the log odds retention of effect hypothesis (3.2) with $\Delta = 0.8$ yields $T = 2.118$ (2.113) using the restricted (unrestricted) estimator for the variance estimation and corresponding p-values 1.71% (1.73%). Thus, we would reject $H_{0,\log\left(\frac{\pi_k}{1-\pi_k}\right)}$ from (2.2) and claim that the test treatment duloxetine is non-inferior over paroxetine.

In contrast to the "classical" hypothesis (2.1), the allocation

$$n_T^* : n_R^* : n_P^* \;=\; 1 : \Delta \, \sqrt{\frac{\pi_{0,T}(1-\pi_{0,T})}{\pi_{0,R}(1-\pi_{0,R})}} : |1-\Delta| \, \sqrt{\frac{\pi_{0,T}(1-\pi_{0,T})}{\pi_{0,P}(1-\pi_{0,P})}} \; . \tag{5.5}$$

is asymptotically optimal. The root terms appear in reciprocal values due to the differences in the variance σ_0^2. It is important to note that the optimal allocation significantly differ due to the formulation of the non-inferiority hypothesis, although the same endpoints of interest are considered.

The major point of the log odds retention of effect hypothesis is the following: unlike for the "classical" hypothesis for binary endpoints and all other hypothesis considered in this work the asymptotically optimal allocation for log odds retention of effect hypothesis, presented in (5.5), does not possess satisfactory finite sample approximation in general, i.e. the optimal allocation for the finite setting significantly differ from the asymptotically optimal one. This issue will be discussed in detail in Section 6 to which we also deferred some exemplary computation of total required sample sizes to obtain an aspired power $1 - \beta$.

Table 5.2: *Precision of sample size formula (4.3) and comparison to the results obtained by Kieser and Friede (2007) for a aspired power of 80% at significance level $\alpha = 2.5\%$.*

$w_T : w_R : w_P$	Δ	$\pi_{0,P}$	$\pi_{0,R}$	Kieser & Friede (2007)		Usage of Eq. (4.3) from this work with exact limit σ_{RML}	
				n	Exact Power	n	Exact Power
1:1:1	0.6	0.1	0.5	309	78.94%	319	80.08%
		0.1	0.7	135	81.51%	132	80.77%
		0.1	0.9	54	83.05%	53	80.49%
		0.3	0.7	318	81.17%	312	80.45%
		0.3	0.9	99	83.92%	94	81.52%
		0.5	0.9	213	84.95%	195	81.43%
	0.8	0.1	0.7	606	81.74%	583	80.18%
		0.1	0.9	201	85.57%	182	81.14%
		0.3	0.9	345	85.39%	309	81.08%
		0.5	0.9	726	84.74%	653	80.51%
2:2:1	0.6	0.1	0.5	270	78.59%	283	80.36%
		0.1	0.7	115	79.96%	119	80.62%
		0.1	0.9	50	84.71%	49	80.71%
		0.3	0.7	290	80.73%	287	80.02%
		0.3	0.9	95	84.25%	89	80.82%
		0.5	0.9	213	86.06%	186	81.11%
	0.8	0.1	0.7	510	81.69%	492	80.15%
		0.1	0.9	170	85.42%	156	81.99%
		0.3	0.9	300	85.51%	269	81.09%
		0.5	0.9	635	84.69%	575	80.88%
3:2:1	0.6	0.1	0.5	252	78.15%	268	80.49%
		0.1	0.7	108	80.54%	110	81.05%
		0.1	0.9	42	80.12%	45	83.09%
		0.3	0.7	276	80.97%	272	80.31%
		0.3	0.9	90	85.70%	83	81.07%
		0.5	0.9	204	87.31%	173	80.65%
	0.8	0.1	0.7	486	82.51%	458	80.17%
		0.1	0.9	156	87.36%	135	81.75%
		0.3	0.9	282	87.17%	241	81.21%
		0.5	0.9	606	86.02%	520	80.30%

5.2   Poisson endpoints: Treatment of epilepsy

In this section we revisit the example in the treatment of epilepsy introduced in Section 2.2.

5.2.1   Performing the RET

The MLE $\hat{\lambda}_k$ is obtained by the mean value $n_k^{-1} \sum_{i=1}^{n_k} X_{ki}$ which is asymptotically normally distributed with variance $\sigma_k^2 = \lambda_k$. The unrestricted MLE of the variance σ^2 is obtained by

$$\hat{\sigma}_{ML}^2 = n \cdot \left(\frac{\hat{\lambda}_T}{n_T} + \Delta^2 \frac{\hat{\lambda}_R}{n_R} + (1-\Delta)^2 \frac{\hat{\lambda}_P}{n_P} \right) .$$

Hence, we end up with the test statistic (see (3.4))

$$T = \frac{-\hat{\lambda}_T + \Delta \hat{\lambda}_R + (1-\Delta)\hat{\lambda}_P}{\sqrt{\frac{\hat{\lambda}_T}{n_T} + \Delta^2 \frac{\hat{\lambda}_R}{n_R} + (1-\Delta)^2 \frac{\hat{\lambda}_P}{n_P}}} \tag{5.6}$$

in order to test $H_{0,-\lambda_k}$ from (2.4), where $H_{0,-\lambda_k}$ is rejected if $T > z_{1-\alpha}$. The restricted version of the Wald-type test is observed by replacing the MLEs $\hat{\lambda}_k$ in the denominator by the to $H_{0,-\lambda_k}$ restricted ones. Again, we have computed the restricted MLEs numerically as for binary endpoints in the previous section.

The RET for the hypothesis (2.4) with $\Delta = 0.5$ yields $T = 1.328$ (1.349) in (5.6) using the restricted (unrestricted) estimator for the variance estimation and corresponding p-values 9.21% (8.86%). Thus, we would not reject $H_{0,-\lambda_k}$ from (2.4) at level $\alpha = 0.05$ and we could not claim that the test treatment is non-inferior over the reference one.

5.2.2   Optimal allocation

Table 5.3: *Optimal allocation of samples for the example of Poisson distributed endpoints*

$\frac{\lambda_{0,T}}{\lambda_{0,P}} = \frac{\lambda_{0,R}}{\lambda_{0,P}}$	$\Delta = 0.5$			$\Delta = 0.7$			$\Delta = 0.8$		
	w_T^*	w_R^*	w_P^*	w_T^*	w_R^*	w_P^*	w_T^*	w_R^*	w_P^*
0.9	0.49	0.25	0.26	0.50	0.35	0.16	0.50	0.40	0.10
0.8	0.49	0.24	0.27	0.49	0.34	0.16	0.49	0.40	0.11
0.7	0.48	0.24	0.28	0.49	0.34	0.17	0.49	0.39	0.12
0.6	0.47	0.23	0.30	0.48	0.34	0.19	0.49	0.39	0.13
0.5	0.45	0.23	0.32	0.47	0.33	0.20	0.48	0.38	0.14
0.3	0.41	0.21	0.38	0.44	0.31	0.24	0.46	0.37	0.17
0.2	0.38	0.19	0.43	0.42	0.30	0.28	0.44	0.36	0.30

For Poisson distributed endpoints and the hypothesis (2.4) the optimal allocation of samples

is given by

$$n_T^* : n_R^* : n_P^* \;=\; 1 : \Delta \sqrt{\frac{\lambda_{0,R}}{\lambda_{0,T}}} : |1 - \Delta| \sqrt{\frac{\lambda_{0,P}}{\lambda_{0,T}}} \;. \tag{5.7}$$

Table 5.3 presents the optimal allocation for the commonly used alternative $\lambda_{0,T} = \lambda_{0,R}$ for different choices of $\lambda_{0,T}/\lambda_{0,P} = \lambda_{0,R}/\lambda_{0,P}$ and Δ. Note that we may assume w.l.o.g $\lambda_{0,P} = 1$ because multiplication of all parameters $\lambda_{0,k}$, $k = T, R, P$, by the same factor does not change the optimal allocation. This simplifies computation significantly. The sample size reductions which are possible are illustrated in Figure 5.2 where the reduction for using the optimal allocation instead of a balanced and a 2:2:1 allocation, respectively, is presented for different values of Δ. The results are quite similar to the ones for binary endpoints in the previous section. For the 2:2:1 allocation we observe reductions between about 5% and 15%. For the balanced allocation reductions between 10% and 20% occur. We again find that the 2:2:1 allocation is more apporiate than the balanced allocation. But again, it is still not the optimal one in considered settings.
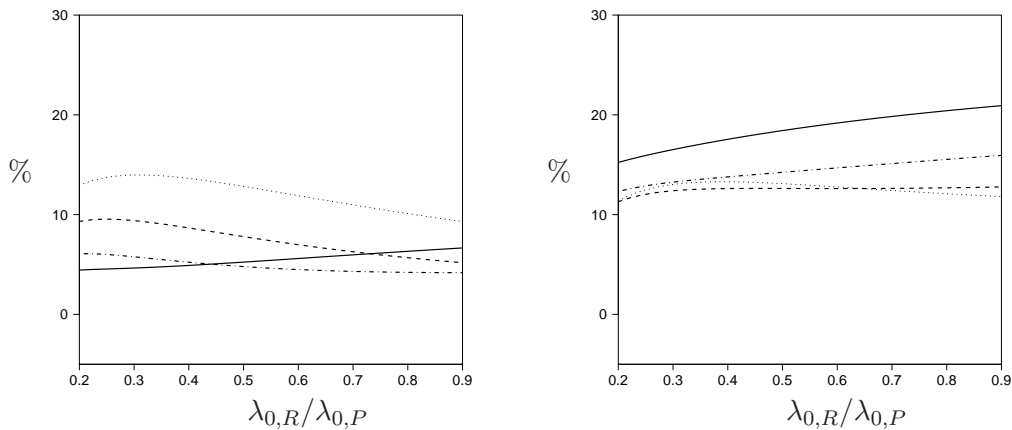

Figure 5.2: *Example of Poisson distributed endpoints:* **Sample size reduction in %** *when optimal allocation is used instead of the balanced allocation (right figure) and instead of the allocation 2:2:1 (left figure) for different values of Δ, $\Delta = 0.5$ (dotted line), $\Delta = 0.6$ (dashed line), $\Delta = 0.7$ (dotdash line), $\Delta = 0.8$ (solid line).*

5.2.3 Planning a trial - applying the GSSP

For Poisson distributed endpoints the weighted KL-divergence is given by

$$K(\zeta^{(0)}, \zeta, w) = \sum_{k=T,R,P} w_k \cdot \left(\lambda_k - \lambda_k^{(0)} + \lambda_k^{(0)} \cdot \left(\log \lambda_k^{(0)} - \log \lambda_k \right) \right) \tag{5.8}$$

with $\zeta = (\lambda_T, \lambda_R, \lambda_P)$ and $\zeta^{(0)} = (\lambda_T^{(0)}, \lambda_R^{(0)}, \lambda_P^{(0)})$. In the following we restrict our investigations to the commonly used alternative $\lambda_T^{(0)} = \lambda_R^{(0)}$. To restrict the minimization problem of the weighted KL-divergence to the boundary of $H_{0,-\lambda_k}$ (2.4) we substitute $\lambda_T = \Delta\lambda_R + (1-\Delta)\lambda_P$ in (5.8). For this situation, an explicit minimization of the KL-divergence is possible. To this end, we evaluate the derivatives of K w.r.t. λ_R and λ_P at zero, which is extremely cumbersome and yields the following rather complex solution for the KL-divergence minimizer, denoted by λ_{k,H_0}, $k = T, R, P$,

$$
\begin{aligned}
\lambda_{R,H_0} \; = \; & \big[\Delta^2(-1+w_T)w_T\lambda_{0,P} - \Delta(-1+w_T)w_T(\lambda_{0,P}-\lambda_{0,T}) + w_R^2((-1+\Delta)\lambda_{0,P}+(2-\Delta)\lambda_{0,T}) \\
& + \; w_R((-1+\Delta)(-1+w_T+\Delta w_T)\lambda_{0,P}+(-\Delta+w_T+2\Delta w_T-\Delta^2 w_T)\lambda_{0,T}) - \mathbf{S}\big] \\
& / \;\; (2(w_R+\Delta(-1+w_T))(w_R+\Delta w_T))
\end{aligned}
$$

$$
\begin{aligned}
\lambda_{P,H_0} \; = \; & \big[w_R^2\lambda_{0,P} + \Delta^2 w_T((-1+w_R+w_T)\lambda_{0,P}-w_R\lambda_{0,T}) + w_R((-1+w_T)\lambda_{0,P}w_T\lambda_{0,T}) \\
& + \; \Delta((2+w_R^2-3w_T+w_T^2+w_R(-3+2w_T))\lambda_{0,P}+(w_R-w_R^2+w_T-w_T^2)\lambda_{0,T}) - \mathbf{S}\big] \\
& / \;\; \big(2((-1+w_R)w_R+\Delta^2(-1+w_T)w_T+\Delta(1-w_T+w_R(-1+2w_T))))
\end{aligned}
$$

$$
\lambda_{T,H_0} \; = \; \Delta\lambda_{R,H_0} + (1-\Delta)\lambda_{P,H_0}
$$

with

$$
\begin{aligned}
\mathbf{S} \; = \; & \big\{-4\Delta(-1+w_R+w_T)((-1+w_R)w_R+\Delta^2(-1+w_T)w_T+\Delta(1-w_T+w_R(-1+2w_T)))\lambda_{0,P} \\
& \cdot \;\; ((-1+w_R+w_T)\lambda_{0,P}-(w_R+w_T)\lambda_{0,T})+(\Delta^2 w_T((-1+w_R+w_T)\lambda_{0,P}-w_R\lambda_{0,T}) \\
& + \;\; w_R((-1+w_R+w_T)\lambda_{0,P}-w_T\lambda_{0,T})+\Delta((2-3w_R+w_R^2-3w_T+2w_Rw_T+w_T^2)\lambda_{0,P} \\
& + \;\; (w_R-w_R^2+w_T-w_T^2)\lambda_{0,T}))^2\big\}^{1/2}.
\end{aligned}
$$

The KL-divergence minimizer λ_{k,H_0}, $k = T, R, P$ are displayed in Table 5.4 (columns 3-5) for different parameter constellations and choices of Δ. Based on these results the limit σ_{RML}^2 of the restricted MLE's of the variance is computed (column 6) and compared to the true variance σ_0^2, see Table 5.4 columns 7 and 8. Throughout Table 5.4 we presumed the usage of the optimal allocation from Table 5.3. In addition, for all parameter constellations the required total samples size $n_{0.7}, n_{0.8}$ to obtain a power of 0.7 and 0.8, respectively, are computed via (4.2) (values in brackets) and (4.3), respectively.

Let us illustrate the use of these tables by the following example. Presume that we would like to detect the alternative $\lambda_T = \lambda_R = 16$, $\lambda_P = 20$ (order of magnitude as in example 2.2) and $\Delta = 0.8$ with a probability of 80% at a significance level $\alpha = 0.05$. Table 5.4 provides that we require a total sample size of 633 under the optimal allocation of $(w_T, w_R, w_P) = (0.49, 0.40, 0.11)$, which yields $n_T = 310$, $n_R = 253$ and $n_P = 70$. A balanced allocation as present in the example data would require a total sample size of 796. Thus, the optimal allocation provides a reduction in the required sample sizes of about 20%.

Table 5.4: **Example of Poisson distributed endpoints:** *Limits of restricted MLE's, limit of variance estimator $\hat{\sigma}_{RML}$ and required samples size to obtain a power of 0.7 and 0.8, respectively, when the variance is estimated restrictedly to the null-hypothesis (unrestrictedly), a nominal significance level $\alpha = 5\%$, for different parameter constellations and choices of Δ for the optimal sample allocation in (5.7).*

Δ	$\frac{\lambda_{0,T}}{\lambda_{0,P}} = \frac{\lambda_{0,R}}{\lambda_{0,P}}$	$\frac{\lambda_{T,H_0}}{\lambda_{0,P}}$	$\frac{\lambda_{R,H_0}}{\lambda_{0,P}}$	$\frac{\lambda_{P,H_0}}{\lambda_{0,P}}$	$\frac{\sigma_{RML}}{\lambda_{0,P}}$	$\frac{\sigma_0}{\lambda_{0,P}}$	$\frac{\sigma_{RML}}{\sigma_0}$	$n_{0.7} \cdot \lambda_{0,P}$	$n_{0.8} \cdot \lambda_{0,P}$
0.5	0.9	0.93	0.88	0.97	1.924	1.923	1.000	6965 (6961)	9150 (9146)
	0.8	0.85	0.75	0.95	1.845	1.842	1.002	1601 (1596)	2102 (2097)
	0.7	0.78	0.64	0.92	1.763	1.755	1.005	649 (645)	852 (847)
	0.6	0.71	0.52	0.89	1.679	1.662	1.011	331 (325)	433 (427)
	0.5	0.64	0.41	0.87	1.594	1.561	1.021	190 (184)	248 (241)
	0.3	0.51	0.21	0.81	1.426	1.322	1.079	76 (68)	98 (89)
	0.2	0.46	0.13	0.80	1.358	1.171	1.160	51 (41)	65 (53)
0.7	0.9	0.92	0.89	0.98	1.913	1.913	1.000	19134 (19131)	25138 (25135)
	0.8	0.83	0.77	0.97	1.822	1.821	1.001	4337 (4333)	5697 (5692)
	0.7	0.75	0.66	0.95	1.726	1.722	1.002	1729 (1724)	2270 (2265)
	0.6	0.66	0.55	0.93	1.624	1.617	1.004	860 (855)	1129 (1123)
	0.5	0.58	0.44	0.91	1.515	1.502	1.009	479 (472)	628 (620)
	0.3	0.42	0.23	0.86	1.278	1.231	1.038	172 (162)	224 (213)
	0.2	0.35	0.14	0.84	1.153	1.060	1.087	105 (92)	135 (121)
0.8	0.9	0.91	0.89	0.99	1.908	1.908	1.000	42813 (42810)	56249 (56245)
	0.8	0.82	0.78	0.98	1.810	1.810	1.000	9640 (9636)	12664 (12660)
	0.7	0.73	0.67	0.97	1.707	1.706	1.001	3810 (3805)	5004 (4999)
	0.6	0.64	0.56	0.95	1.597	1.594	1.002	1875 (1869)	2462 (2456)
	0.5	0.55	0.46	0.94	1.479	1.473	1.004	1028 (1021)	1349 (1342)
	0.3	0.38	0.25	0.90	1.210	1.186	1.020	348 (338)	456 (444)
	0.2	0.30	0.15	0.87	1.054	1.005	1.049	200 (186)	261 (244)

5.3   Censored, exponentially distributed endpoints

In this section we summarize the paper Mielke et al. (2008), which was already citied several times in this work and which has formed the impulse for the general theory presented in the Chapters 3 and 4. Mielke et al. (2008) consider the assessment of non-inferiority for censored, exponentially distributed endpoints by means of a retention of effect type hypothesis. As exponentially distributed endpoints fit certainly in the general parametric model considered in Chapter 3 and 4, this setting just represents a specific application of presented results. However, Mielke et al. (2008) allow for right-censoring of the observations, which therewith extends the results so far. The work was mainly motivated by a clinical trial in depression, which we would like to introduce in the next section.

5.3.1   Refined investigation of the treatment in depression
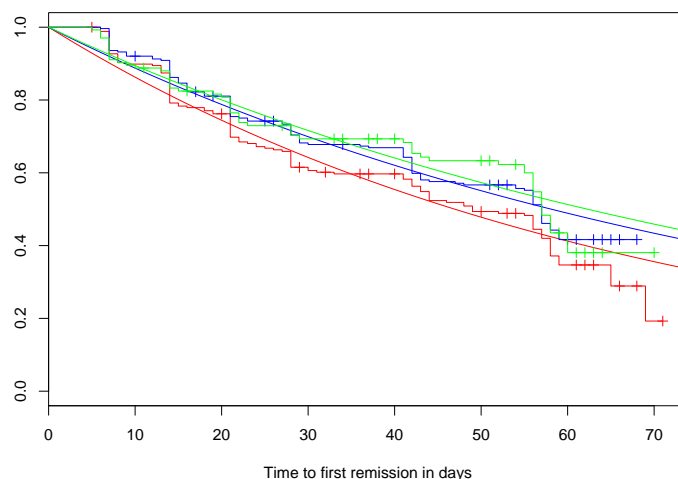
Figure 5.3: *Three-arm study for treatment of depression: Kaplan-Meier curves with marks for censored data and fitted exponential survival curves for the endpoint "time to first response", test treatment (red), reference treatment (blue) and placebo (green).*

In the therapy of depression achieving remission is the clinically desired goal (Nierenberg and Wright, 1999), whereas the remission is defined as maintaining the Hamilton Rating Scale of Depression (HAM-D) total score at ≤ 7. The examples considered in this work (confer Section 2.1 and 5.1) and Kieser and Friede (2007) provide the statistical methodology to examine remission as binary endpoint or to be more precisely, does the patient achieve remission at treatment end or not. However, Yadid et al. (2000) point out that in addition to remission the fast onset of action and the prevention of relapse are important and thus are the major

goal of the present research. The primary endpoint time to first remission incorporates this issue. The occurrence of remission can be investigated over the complete time interval of the study. The data representing the time to first remission in Figure 5.3 are from a randomized, double-blind study in major depression, where a new antidepressant is compared to a standard antidepressant known for having a fast onset of action and to placebo. We assume that the time points to first remission are i.i.d. right censored, exponential distributed in each group. The PP-plots in Figure 5.4 indicate a quite good fit of this model. Note that due to the heavy censoring at the right tail the quality of fit is decreased, of course.
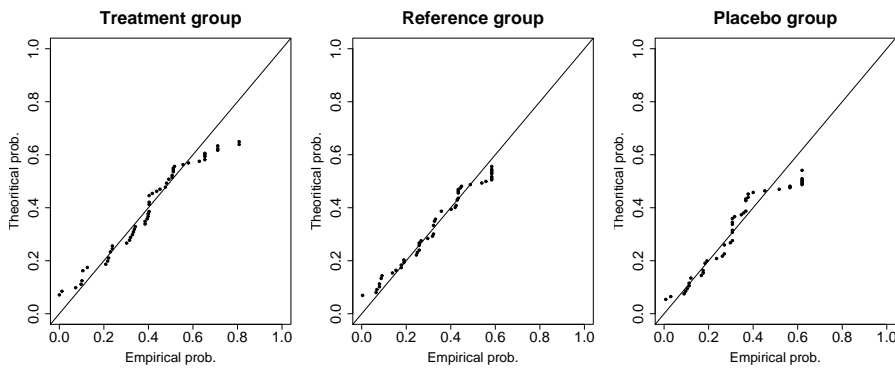
Figure 5.4: *PP-Plots for fitted exponential model vs. Kaplan-Meier.*

5.3.2 Model and hypothesis

The model considered by Mielke et al. (2008) is formalized in following way. Let T_{ki} for $i = 1, \ldots, n_k$ be independent and exponentially distributed survival times with parameter λ_k, $k = R, T, P$. Due to the interpretation of the parameter and therewith combined formulation of the hypothesis it is assumed that the parametrization of the exponential distribution is such that $\mathrm{E}[T_{ki}] = \lambda_k$. Further, let the corresponding censoring times U_{ki} be independent distributed of T_{ki} for $i = 1, \ldots, n_k$ and $k = R, T, P$. The observations consist of pairs (X_{ki}, δ_{ki}), where $X_{ki} = \min\{T_{ki}, U_{ki}\}$ are the observed survival times and $\delta_{ki} = \mathbf{1}_{\{T_{ki} \leq U_{ki}\}}$, $i = 1, \ldots, n_k$, $k = R, T, P$, are the corresponding censoring indicators. Hence $\delta_{ki} = 1$ stands for an uncensored observation. Further, it is assumed that the probabilities for an uncensored observation should be positive, i.e.

$$p_k := P(\delta_{ki} = 1) > 0$$

for $k = R, T, P$. Thus, the distribution of the observations (X_{ki}, δ_{ki}) are characterized by the parameter $\theta_k = (\lambda_k, p_k)$. According to the example in the treatment of depression above it is assumed that small values for the observations X_{ki} are associated with higher efficacy of the

treatment, e.g. we observe the time which elapses until healing or in general until a positive impact occurs. Therefore, small values of λ_T are desirable. The hazard ratio, which is in the case of exponentially distributed endpoints just the ratio of the λ's, is the usual way of comparing time to event endpoints. Therefore, Mielke et al. (2008) consider as measure of efficacy $h(\lambda_k, p_k) = -\log \lambda_k$ yielding the retention of effect hypothesis

$$\text{vs.} \quad \begin{aligned} H_{0,-\log \lambda_k} &: \ \log \lambda_T - \log \lambda_P \geq \Delta \left(\log \lambda_R - \log \lambda_P\right) \\ K_{0,-\log \lambda_k} &: \ \log \lambda_T - \log \lambda_P < \Delta \left(\log \lambda_R - \log \lambda_P\right) \end{aligned} \tag{5.9}$$

with $\Delta \in [0, \infty)$. Alternatively the retention of effect hypothesis from (5.9) could be defined in the ratio of differences in means, i.e. through $H_{0,-\lambda_k} : (\lambda_P - \lambda_T) \leq \Delta(\lambda_P - \lambda_R)$, (Hung et al., 2005). The methods presented in this work also provide basics for a Wald-type test for $H_{0,-\lambda_k}$. However, in this case the asymptotic variance would depend on the parameters λ_k, $k = R, T, P$. In contrast for the RET for the hypothesis (5.9) the asymptotic variance σ^2 is independent of the parameters λ_k, $k = R, T, P$. That has the advantage that the variance can be estimated unrestricted, see (5.10).

5.3.3   RET and optimal sample allocation

According to the general procedure summarized in Figure 3.1 in Chapter 3 Mielke et al. (2008) obtain for $H_{0,-\log \lambda_k}$ in (5.9) as a test statistic

$$T = \sqrt{n} \, \frac{\hat{\eta}}{\hat{\sigma}} = \frac{\log \hat{\lambda}_T - \Delta \log \hat{\lambda}_R + (\Delta - 1)\log \hat{\lambda}_P}{\sqrt{\frac{1}{\delta_T} + \frac{\Delta^2}{\delta_R} + \frac{(\Delta-1)^2}{\delta_P}}} \, , \tag{5.10}$$

with the MLE for λ_k

$$\hat{\lambda}_k = \frac{X_k}{\delta_k} \, ,$$

for $k = R, T, P$ with $X_k = \sum_{i=1}^{n_k} X_{ki}$ and $\delta_k = \sum_{i=1}^{n_k} \delta_{ki}$. As usual, for a given level of significance α the hypothesis $H_{0,-\log \lambda_k}$ will be rejected and non-inferiority can be claimed if $T < z_\alpha$, where z_α denotes the α-quantile of the standard normal distribution.

Consistent with the optimal allocation derived in Chapter 4, Figure 4.1, Mielke et al. (2008) obtain that the optimal sample allocation is given by

$$n_T^* : n_R^* : n_P^* = 1 : \Delta\sqrt{p_T/p_R} : |1-\Delta|\sqrt{p_T/p_P},$$

which yields a minimal total required sample size of

$$n_{1-\beta} = \left(\frac{1}{\sqrt{p_T}} + \frac{\Delta}{\sqrt{p_R}} + \frac{|1-\Delta|}{\sqrt{p_P}}\right)^2 \left(\frac{z_\alpha - z_{1-\beta}}{\eta^{(0)}}\right)^2 . \tag{5.11}$$

From (5.11) we see that the total required sample size is a monotone decreasing function in each p_k, $k = T, R, P$, and it is minimal in the case of uncensored observations, i.e. $p_T = p_R = p_P = 1$. Further, the monotonicity provides a *worst case scenario* for sample size planning by means of presuming homogeneous censoring probabilities in the three groups, $k = T, R, P$, and setting the common censoring probability to the smallest value, i.e. $p = \min\{p_T, p_R, p_P\}$. In planning a clinical trial, one would expect $p_T, p_R > p_P$ and hence $p_P = \min\{p_T, p_R, p_P\}$ because the reference and the test treatment are expected to be efficient, i.e. $\lambda_T, \lambda_R < \lambda_P$, which implies under identically censoring variables U_k in the groups that reference and test treatment are less affected by censoring than placebo. Hence, a conservative recommendation for planning the trial is to assume that all censoring probabilities equal p_P.
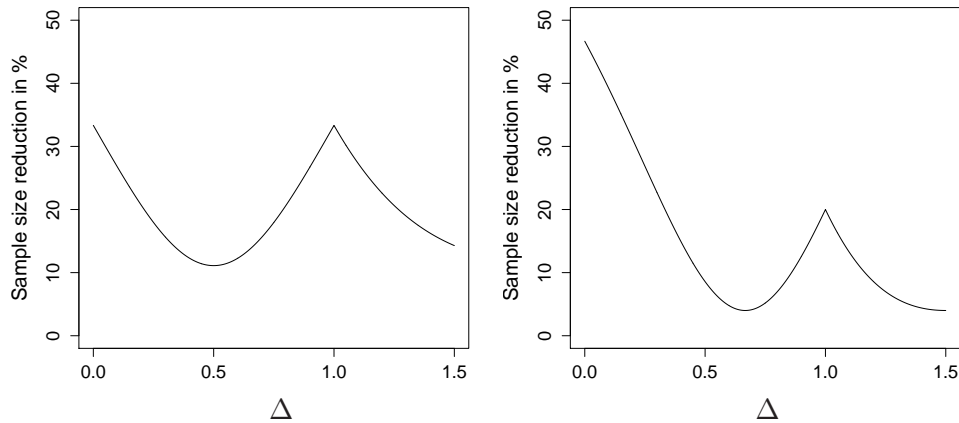
Figure 5.5: *Reduction in total sample size when optimal allocation is used instead of balanced allocation (left figure), and instead of 2:2:1 allocation (right figure).*

As in the examples of binary and Poisson endpoints from the precedent sections the optimal allocation partly permits a serious reduction in required total sample sizes. This is illustrated in Figure 5.5, where the reduction for using the optimal allocation instead of a balanced and $2 : 2 : 1$ allocation, respectively, is presented. For the balanced design a reduction of at least 10% is always possible and even more than 30% for Δ close to zero or one. The allocation $2 : 2 : 1$ is more appropriate for $\Delta \in [0.5, 1)$, but a reduction up to 20% is still possible by reallocating to the optimal allocation.

5.3.4 Complete test procedure

As already mentioned in the introduction, Chapter 1, the test problem considered so far is to show non-inferiority of the test treatment over the reference. The inclusion of a placebo group makes it possible to directly demonstrate the effectiveness of a therapy and therewith

ensures assay sensitivity of the test procedure. Pigeot et al. (2003) carry out a pretest for superiority of the reference treatment over placebo, which provides internal evidence of assay sensitivity. Though, Koch (2005) points out that this procedure would blame a test treatment that has shown to be superior over placebo and non-inferior over the reference for the fact that reference could not beat placebo. Therefore, Koch and Röhmel (2004) perform a pretest for superiority of the test treatment over placebo instead. It was not the objective of Mielke et al. (2008) to take up this discussion. Thus, they only notice that in any case a two-step test procedure must be conducted to establish non-inferiority and effectiveness of the test treatment, where in a first step a pretest for superiority of either the reference or the test treatment over placebo is performed, and in a second step the non-inferiority is investigated. For the specific setting of censored, exponentially distributed endpoints from Mielke et al. (2008) the pretest for superiority of a treatment over placebo coincides with rejecting the null hypothesis $H_{S,j} : \lambda_j \geq \lambda_P$ either for the reference treatment ($j = R$) or for the test treatment ($j = T$). Thus, the overall hypothesis is given by

$$H_0 : \quad H_{0,-\log \lambda_k} \cup H_{S,j} \; = \; \left\{\log \lambda_T - \log \lambda_P \geq \Delta \left(\log \lambda_R - \log \lambda_P\right)\right\} \cup \left\{\lambda_j \geq \lambda_P\right\},$$

where H_0 is rejected if the sub-tests for $H_{0,-\log \lambda_k}$ and $H_{S,j}$ can be rejected. In order to avoid a misunderstanding, note that H_0 either includes $H_{S,R}$ or $H_{S,T}$ and not both at once. Due to the principles of intersection-union-tests, this test decision for H_0 does not exceed a level α if $H_{0,-\log \lambda_k}$ and $H_{S,j}$ are tested the level α, respectively. Therefore, the power for rejecting H_0 is reduced compared to simple testing $H_{0,-\log \lambda_k}$. However, Mielke et al. (2008) figure out that this reduction is negligible for the commonly used alternative $\lambda_T = \lambda_R < \lambda_P$, see Section 4.2. in Mielke et al. (2008). Similar results were obtained by Pigeot et al. (2003) and Kieser and Friede (2007) for normal and binomial endpoints, respectively.

We would like to note that Mielke et al. (2008) figure out that in their setting from a statistical point of view the pretest for $H_{S,T}$ is preferred to those for $H_{S,R}$ because the test for superiority over placebo is more powerful for the test treatment than for the reference treatment in general (for the settings considered by Mielke et al. (2008)), but especially also for the commonly used alternative $\lambda_T = \lambda_R$ due to optimal sample allocation which gives more weight to the test treatment group.

5.3.5 Clinical trial in depression

We revisit the example in treatment of major depression from Section 5.3.1. In this randomized, double blind study a new antidepressant (T) is compared to a standard antidepressant (R), known for having a fast onset of action, and to placebo (P). The comparison is based on the analysis of the time to first remission whereas remission is defined as maintaining the

Table 5.5: *P-values for the RET and the pretests.*

Hypothesis	p-value in %
$H_{0,-\log \lambda_k}$ with $\Delta = 0.5$	1.83
$H_{0,-\log \lambda_k}$ with $\Delta = 0.8$	2.51
$H_{0,-\log \lambda_k}$ with $\Delta = 1$	4.42
$H_{S,R}$	33.34
$H_{S,T}$	3.88

Hamilton Rating Scale of Depression (HAM-D) total score at ≤ 7 as aforementioned. The data set consists of $n_T = 262$, $n_R = 267$, and $n_P = 135$ pairs of observations, the time to first remission in days, and the censoring indicator with a fraction of 0.51, 0.46, and 0.41 uncensored observations, respectively. For the ML-estimators we obtain $\hat{\lambda}_T = 67.75$, $\hat{\lambda}_R = 83.84$, and $\hat{\lambda}_P = 89.87$. Thus, one would guess that the new antidepressant has the fastest onset of action followed by the reference treatment and by placebo.

The resulting p-values for the RET and the pretests are presented in Table 5.5. If we presume the commonly used significance level of 5%, the hypothesis of the RET could be even rejected for $\Delta = 1$ and hence not only non-inferiority but also superiority of the new treatment over the standard treatment could be claimed. The pretest with the new treatment ($H_{S,T}$) would reject in favor of superiority of new treatment over placebo. In contrast, the pretest with the reference treatment ($H_{S,R}$) would fail, i.e. it does not reject. This fact supports the view of Koch and Röhmel (2004) to perform the pretest for $H_{S,T}$ instead of $H_{S,R}$.

The present sample size allocation is approximately $n_T : n_R : n_P \approx 2 : 2 : 1$. Hence, if we consider $\Delta = 0.5$, a sample size reduction of roughly 10% would have been possible by reallocating to the optimal allocation $2 : 1 : 1$ (see Figure 5.5).

FINITE SAMPLE APPROXIMATIONS

The procedures presented and applied in the previous sections, including performing the RET, the sample size formulas and the optimal allocation of samples, make use of the asymptotic normality of ML-estimators, i.e. they are based on results which are valid for the number of total samples n converging to infinity. In Section 6.1 we discuss the finite sample behavior of this normal approximation. For the derivation of the asymptotically optimal allocation in Section 4.1 the asymptotically subordinated terms were disregarded, which for this reason yields to a further approximation beside the normal approximation. In Section 6.2 we investigate the finite sample effect of this procedure.

6.1 Normal approximation

The test decision for the RET in (3.4) as well as the sample size formulas (4.2) and (4.3) for planning the RET make use of the normal approximation of the test statistic T in (3.4). As in this work we are faced with regular parametric models it is to expect that the normal approximation fits well for quite moderate sample sizes. Indeed, extensive simulations studies by Kieser and Friede (2007) and Mielke et al. (2008) confirm for normally and exponentially distributed endpoints, respectively, that the approximation works quite well for $n > 100$, i.e. the actual type I error nearly attains the nominal significance level exactly. However, note that for binary endpoints restricted estimation of the variance is necessary to obtain this improved approximation results due to the dependency of the variance on the parameters of interest. For Poisson distributed endpoints we have obtained similar results, which are omitted for the reason of conciseness. We only exemplary present the results for censored, exponentially distributed endpoints. Figure 6.1 displays the simulated actual type I errors with 10000 replications for the balanced, the $2 : 2 : 1$ and the optimal allocation, respectively, for $\Delta = 0.5$. The results for $\Delta = 0.7$ are similar and omitted again due to the marginal gain

of insight.

Table 6.1: *Simulated actual type I error in % for a nominal significance level of $\alpha = 0.05$, $p_T = p_R = p_P = 0.8$, $\Delta = 0.5$, and $10\,000$ replications. Values **larger than 5.5%** are bold.*

n	$n_T : n_R : n_P$	λ_P/λ_R						
		10	8	5	3	2	1.5	1.2
30	1:1:1	**5.69**	**5.89**	**5.71**	**5.71**	**6.02**	**5.94**	**6.17**
	2:2:1	5.35	5.10	5.31	5.26	4.91	5.02	5.20
	2:1:1	4.68	4.44	4.55	4.78	4.64	4.78	4.63
60	1:1:1	**5.90**	**5.67**	**5.62**	5.29	5.21	**5.59**	5.35
	2:2:1	5.23	5.25	4.91	4.56	5.22	5.19	5.26
	2:1:1	4.92	4.23	4.79	4.87	4.84	4.32	5.02
120	1:1:1	**5.80**	5.29	5.47	5.35	**5.60**	5.42	5.47
	2:2:1	4.83	5.00	5.28	4.80	4.88	5.35	4.66
	2:1:1	4.76	4.46	4.30	4.94	4.91	4.56	4.73
240	1:1:1	4.97	5.31	5.43	5.22	4.96	5.09	5.30
	2:2:1	5.05	5.20	5.03	5.06	4.87	5.11	4.87
	2:1:1	5.02	4.77	4.54	5.05	4.79	5.06	4.61
480	1:1:1	5.01	**5.68**	5.11	5.03	5.01	5.10	5.15
	2:2:1	5.13	5.41	5.06	4.86	5.19	5.32	5.07
	2:1:1	4.90	4.75	5.15	5.13	4.55	4.94	5.00
960	1:1:1	5.35	5.13	4.98	5.42	5.02	5.22	5.09
	2:2:1	4.89	4.77	4.74	5.14	4.87	4.83	4.97
	2:1:1	4.61	4.89	4.74	4.63	4.98	5.09	4.64
1440	1:1:1	4.94	5.23	5.08	5.37	4.98	5.25	5.31
	2:2:1	4.84	4.95	5.25	5.08	4.83	4.83	4.75
	2:1:1	5.03	5.10	5.23	5.27	4.60	4.89	4.63

6.2   Asymptotically optimal sample allocation

The asymptotically optimal allocation in (4.1) has to be understood as approximation for finite samples as it customary for asymptotic results. In Section 4.1.2 we have shown that maximizing the power asymptotically is tantamount to minimize the variance σ_0^2. However, if the variance σ^2 is estimated with restriction to the null hypothesis the term σ_{RML} in the power formula (3.12), which is also influenced by the allocation, is disregarded. Asymptotically this is correct because the term σ_0^2 is dominating due to $\sqrt{n}$ factor. Nevertheless, as we will see in the following this may affect the optimal allocation for finite sample sizes.

It is important to note that we only disregard the term σ_{RML} when we estimate the variance σ_0^2 with restriction to the null hypothesis. Thus, the case of normal endpoints from Pigeot et al. (2003) and the case of censored, exponentially distributed endpoints from Mielke et al. (2008) are not faced with this approximation and possible resulting inaccuracies because in these settings the variance σ_0^2 does not depend on the parameters of interest. For binary
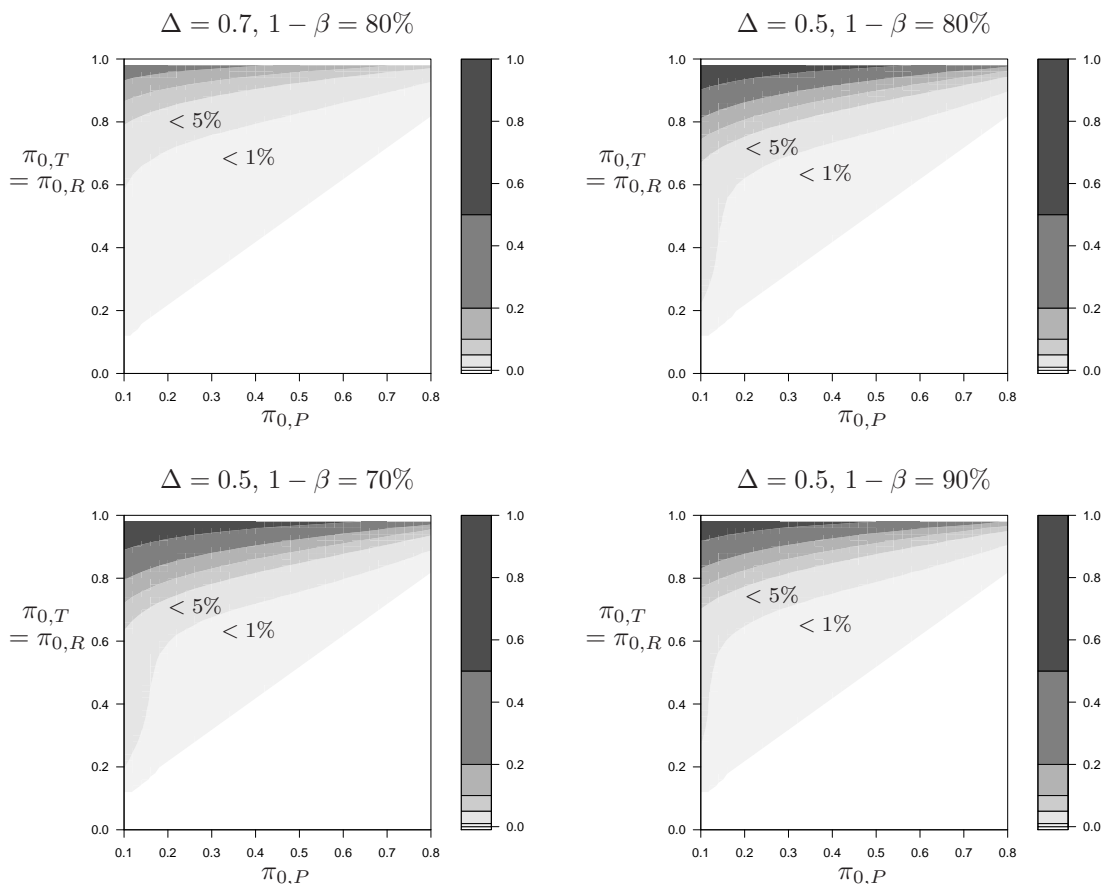
Figure 6.1: *Finite sample reduction from (6.2) of minimal required total sample size relative to the* **asymptotic optimal allocation***.*

endpoints with the "classical" hypothesis (2.1) and for Poisson endpoints with the hypothesis considered in this work (2.4) numerically investigations turn out that the asymptotically optimal allocation w_T^*, w_R^*, w_P^* from (4.1) performs satisfactory for finite samples, in general, in terms of that no other allocation exhibits a significant lower total number of required sample size. From the examples considered in this work binary endpoints with the log odds retention of effect hypothesis (2.2) take a particular position in this context. In this setting the asymptotically optimal allocation is even not nearly optimal (meaning is given in the next section) for finite samples, in general. Hence, we focus in the following on this setting and discuss this setting in the next section in detail.

6.2.1   The log odds RET

As already discussed the asymptotically optimal allocation (5.5) for the log odds retention of effect hypothesis and binary endpoints has to be understood as approximative for finite sam-
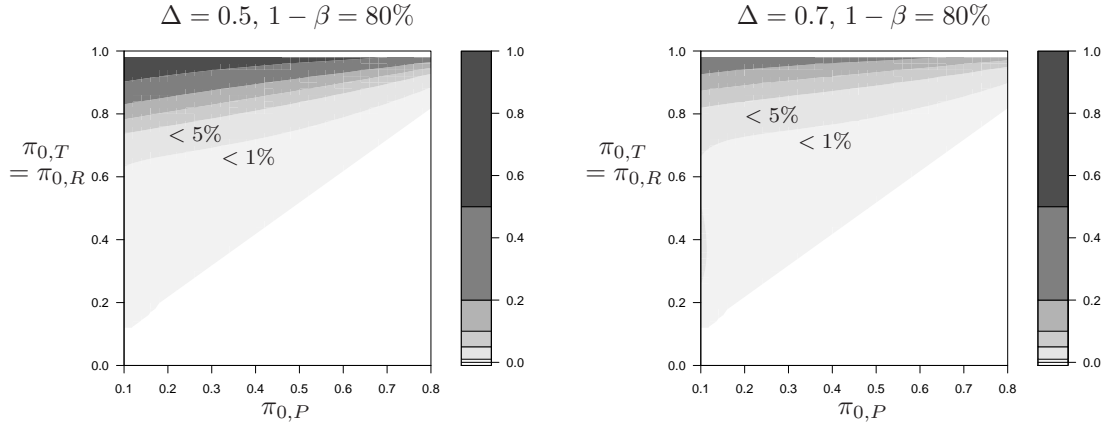
Figure 6.2: *Finite sample reduction from (6.2) of minimal required total sample size relative to the **simplified optimal allocation*** $1 : \Delta : |1 - \Delta|$.

ples. To investigate the finite sample behavior we compare in this section the asymptotically optimal allocation to the *finite sample optimal allocation*, denoted by w_T^f, w_R^f, w_P^f.

To this end, we compute the finite sample optimal allocation numerically, i.e. minimizing the total required sample size from (4.3)

$$n_{1-\beta}(w_T, w_R) \;=\; (z_{1-\alpha} \cdot \sigma_{RML} + z_{1-\beta} \cdot \sigma_0)^2 \cdot \eta_0^{-2} \tag{6.1}$$

in w_T, w_R. Note that we write $n_{1-\beta}(w_T, w_R)$ as a function of w_T, w_R, w_P, now, to highlight the dependency on the allocation, in contrast to (4.3). w_P is omitted because it is determined through $w_P = 1 - w_T - w_R$. The minimization of $n_{1-\beta}$ is performed by means of a grid search with $w_T = 0.01, \ldots, 0.98$ and $w_R = 0.01, \ldots, (0.99 - w_T)$.

It is self-evident that the allocations should be compared with respect to their resulting required sample sizes. To this end, we consider the reduction in total sample sizes for the finite optimal allocation relative to the asymptotic optimal allocation

$$R = \frac{n_{1-\beta}(w_T^*, w_R^*) - n_{1-\beta}(w_T^f, w_R^f)}{n_{1-\beta}(w_T^*, w_R^*)} \tag{6.2}$$

with $w_T^* = n_T^*/n$ and $w_R^* = n_R^*/n$ from (5.5). Note that the asymptotically optimal allocation is also optimal in the finite setting if the reduction R in (6.2) is zero. We have performed the comparison for many settings which we do not discuss in detail. The results for the commonly used settings of $\Delta = 0.5$ and $\Delta = 0.7$ are presented in Figure 6.1. The reduction in total sample size from (6.2) is plotted in dependency of $\pi_{0,P}$ and $\pi_{0,T} = \pi_{0,R}$ for a significance level of $\alpha = 5\%$ and a power of $1 - \beta = 80\%$. To illustrate the influence of the power $1 - \beta$ we have included for $\Delta = 0.5$ plots for $1 - \beta = 70\%, 90\%$, in addition. It turns out that the asymptotically optimal allocation is also nearly optimal in the finite setting for a broad field of

parameters $\pi_{0,P}$ and $\pi_{0,T} = \pi_{0,R}$, more precisely as long as $\pi_{0,T} = \pi_{0,R}$ is not very much larger than $\pi_{0,P}$. To this end, we classify an allocation as nearly optimal when the reduction in (6.2) by the optimal allocation is less than 5%. However, the asymptotically optimal allocation is no more optimal if $\pi_{0,T} = \pi_{0,R}$ is very large compared to $\pi_{0,P}$, e.g. if $\pi_{0,P} = 0.5$ and $\pi_{0,T} = \pi_{0,R} > 0.92$ or $\pi_{0,P} = 0.8$ and $\pi_{0,T} = \pi_{0,R} > 0.98$ for $\Delta = 0.7$ and $1 - \beta = 80\%$ (Figure 6.1, upper left). With decreasing Δ the area, where the asymptotically optimal allocation is even not nearly optimal, increases, e.g. for $\Delta = 0.5$ and $1 - \beta = 80\%$ (Figure 6.1, upper right) the asymptotically optimal allocation is not nearly optimal in the finite setting for $\pi_{0,P} = 0.5$ and $\pi_{0,T} = \pi_{0,R} > 0.86$ or $\pi_{0,P} = 0.8$ and $\pi_{0,T} = \pi_{0,R} > 0.94$. As to be expected, the area, where the asymptotically optimal allocation is even not nearly optimal, decreases with increasing power $1 - \beta$, confer Figure 6.1, upper right, lower left and right plots, where the reduction in total sample size is displayed for $\Delta = 0.5$ and $1 - \beta = 70\%, 80\%$ and 90%, respectively.

Practical recommendation: The asymptotically optimal allocation (5.5) depends on the choice of alternative $(\pi_{0,T}, \pi_{0,R}, \pi_{0,P})$. In most practical situations the specification of one single alternative is undesirable or even impossible, e.g. one wants to consider more than one alternative. For this, we recommend to use the *simplified optimal allocation* $1 : \Delta : |1 - \Delta|$ as a rule of thumb. We have evaluated the finite sample performance analogously to the asymptotically optimal one. The results are presented in Figure 6.2 for $\alpha = 5\%$, $1 - \beta = 80\%$, $\Delta = 0.5$ and 0.7, respectively. The results for the asymptotically optimal and the simplified optimal allocation are quite the same, i.e. the simplified optimal allocation also performs well where the asymptotically optimal allocation does, i.e. the asymptotically optimal allocation $w_T^* : w_R^* : w_P^*$ provides no significant improvement relative to the simplified optimal allocation $1 : \Delta : |1 - \Delta|$, when the variance is estimated with restriction to the null hypothesis.

Therefore, in summary, we recommend for the general proceeding in sample size planning to use the simplified optimal allocation

$$w_T^s : w_R^s : w_P^s = 1 : \Delta : |1 - \Delta| \tag{6.3}$$

where no significant further finite total sample reduction is possible (reduction $< 5\%$, can be determined from Figure 6.2) and apart from that compute the finite sample optimal allocation numerically.

The latter case is illustrated in Table 5.1. Here, the total required sample sizes $n_{0.8}(\cdot, \cdot)$ from (6.1) to obtain a power of 0.8 for the asymptotically optimal (5.5), the simplified optimal (6.3) and the finite sample optimal allocation is displayed. This is done for different parameter constellations in the alternative, $\pi_{0,P}$ and $\pi_{0,T} = \pi_{0,R}$, for $\Delta = 0.5$ and a significance level $\alpha = 5\%$. In addition, the underlying asymptotically optimal allocation $w_T^* : w_R^* : w_P^*$ and the finite sam-

ple optimal allocation $w_T^f : w_R^f : w_P^f$ are presented for each parameter constellation. To this end, the finite sample optimal allocation is computed by means of a grid search as above. According to our recommendation the total required sample size nearly coincide for the asymptotically and the simplfied optimal allocation as long as no significant further reduction is possible by the finite sample optimal allocation. In settings where the asymptotically and the simplified optimal allocation, respectively, lack of satisfactory finite sample approximations (i.e. $\pi_{0,P} \ll \pi_{0,T} = \pi_{0,R}$) we observe the following rule of thumb: starting from asymptotically optimal allocation more weight is given to the reference group relative to the test group to reduce the total required sample sizes in the finite setting, whereas the weight of the placebo group remains nearly unchanged. Consider for example $\pi_{0,P} = 0.5$ and $\pi_{0,T} = \pi_{0,R} = 0.96$ with asymptotically optimal allocation $w_T^* : w_R^* : w_P^* = 0.590 : 0.295 : 0.116$ and finite sample optimal allocation $w_T^f : w_R^f : w_P^f = 0.37 : 0.52 : 0.11$. Here, the fractions of the reference and the test group are nearly interchanged and the placebo group fraction remains unchanged. Note that in this setting the finite sample optimal allocation reduces the total required sample size about 50%.

Table 6.2: **Required total sample sizes**: *Required samples size $n_{0.8}(\cdot,\cdot)$ to obtain a power of 0.8 for different parameter constellations, $\Delta = 0.5$ and significance level $\alpha = 5\%$ under different allocation, the asymptotically optimal sample allocation $w_T^* : w_R^* : w_P^*$, the simplified optimal allocation $w_T^s : w_R^s : w_P^s = 1 : \Delta : |1 - \Delta| = 1 : 0.5 : 0.5$ and the finite sample optimal allocation $w_T^f : w_R^f : w_P^f$.*

$\pi_{0,P}$	$\pi_{0,T}=\pi_{0,R}$	w_T^*	w_R^*	w_P^*	w_T^f	w_R^f	w_P^f	$n_{0.8}(w_T^*,w_R^*)$	$n_{0.8}(w_T^s,w_R^s)$	$n_{0.8}(w_T^f,w_R^f)$
0.1	0.3	0.442	0.221	0.337	0.50	0.21	0.29	321	323	318
	0.5	0.429	0.214	0.357	0.48	0.24	0.28	103	105	102
	0.7	0.442	0.221	0.337	0.43	0.31	0.26	60	62	58
	0.8	0.462	0.231	0.308	0.40	0.37	0.23	58	61	51
	0.9	0.500	0.250	0.250	0.34	0.49	0.17	103	103	53
	0.94	0.527	0.264	0.209	0.30	0.57	0.13	223	198	63
	0.96	0.547	0.274	0.179	0.27	0.63	0.10	448	365	76
0.2	0.5	0.471	0.235	0.294	0.49	0.25	0.26	227	227	226
	0.7	0.482	0.241	0.276	0.46	0.30	0.24	98	99	97
	0.8	0.500	0.250	0.250	0.43	0.36	0.21	84	84	76
	0.94	0.557	0.278	0.165	0.33	0.55	0.12	209	174	82
	0.96	0.573	0.287	0.140	0.30	0.60	0.10	372	287	97
0.3	0.5	0.489	0.244	0.267	0.50	0.25	0.25	572	572	572
	0.7	0.500	0.250	0.250	0.47	0.29	0.24	162	162	160
	0.8	0.516	0.258	0.225	0.44	0.35	0.21	120	118	112
	0.9	0.547	0.274	0.179	0.39	0.45	0.16	135	123	96
	0.94	0.568	0.284	0.147	0.35	0.52	0.13	211	176	104
	0.96	0.583	0.292	0.125	0.32	0.58	0.10	341	265	119
0.4	0.7	0.508	0.254	0.238	0.48	0.28	0.24	288	288	287
	0.8	0.524	0.262	0.214	0.46	0.33	0.21	179	177	171
	0.9	0.554	0.277	0.170	0.41	0.42	0.17	165	152	129
	0.94	0.574	0.287	0.139	0.37	0.50	0.13	225	190	131
	0.96	0.588	0.294	0.118	0.35	0.55	0.10	334	265	145
0.5	0.7	0.511	0.255	0.234	0.49	0.27	0.24	627	626	625
	0.8	0.526	0.263	0.211	0.47	0.31	0.22	293	291	286
	0.9	0.556	0.278	0.167	0.43	0.40	0.17	213	200	180
	0.94	0.576	0.288	0.137	0.39	0.47	0.14	254	221	171
	0.96	0.590	0.295	0.116	0.37	0.52	0.11	345	281	182
0.6	0.8	0.524	0.262	0.214	0.48	0.30	0.22	587	584	580
	0.9	0.554	0.277	0.170	0.45	0.37	0.18	304	291	273
	0.94	0.574	0.287	0.139	0.41	0.44	0.15	310	279	236
	0.96	0.588	0.294	0.118	0.39	0.49	0.12	381	321	237
0.7	0.8	0.516	0.258	0.225	0.49	0.28	0.23	1995	1993	1988
	0.9	0.547	0.274	0.179	0.46	0.34	0.20	523	511	495
	0.94	0.568	0.284	0.147	0.44	0.40	0.16	432	402	365
	0.96	0.583	0.292	0.125	0.41	0.45	0.14	466	409	339
0.8	0.9	0.533	0.267	0.200	0.48	0.30	0.22	1482	1472	1456
	0.94	0.557	0.278	0.165	0.46	0.36	0.18	805	778	744
	0.96	0.573	0.287	0.140	0.44	0.40	0.16	705	654	595

Iᴍᴘʟᴇᴍᴇɴᴛᴀᴛɪᴏɴꜱ

In the foregoing chapters we have discussed the planning and analyzing of the RET in general and for several specific endpoints in detail. Throughout the following tasks are considered: (a) Performing the RET (b) optimal sample allocation and (c) sample size formulas. According to this we provide in the Appendix A the R source code of functions including detailed documentation for performing this tasks, i.e. planning and analyzing non-inferiority trials with retention of effect hypothesis, for all endpoints discussed and mentioned in this work. This covers binary (Section 2.1 and 5.1), Poisson (Section 2.2 and 5.2), normally (Pigeot et al., 2003) and censored, exponentially distributed endpoints (Mielke et al., 2008, and Section 5.3). All provided functions have the following common structure:

RET.xx.yy( )	Performs the RET for given data
RET.xx.yy.OptAlloc( )	Computes the optimal sample allocation for the RET
RET.xx.yy.Samplesize( )	Determines the required sample sizes for the RET

where 'xx' specifies the distribution of the endpoints and 'yy' the retention of effect hypothesis.

BINARY ENDPOINTS with "classical" hypothesis (Appendix A.1.1):

Model: $X_{ki} \sim B(\pi_k)$ for $k = T, R, P$ and $i = 1, \ldots, n_k$.

Hypothesis: $H_{0,\pi_k} : \pi_T - \pi_P \leq \Delta \cdot (\pi_R - \pi_P)$

Functions: *RET.Bin.Id()*

RET.Bin.Id.OptAlloc()

RET.Bin.Id.Samplesize()

BINARY ENDPOINTS with log odds hypothesis (Appendix A.1.2):

Model: $X_{ki} \sim B(\pi_k)$ for $k = T, R, P$ and $i = 1, \ldots, n_k$.

Hypothesis: $H_{0,\log\left(\frac{\pi_k}{1-\pi_k}\right)} : \log\left(\frac{\pi_T}{1-\pi_T}\right) - \log\left(\frac{\pi_P}{1-\pi_P}\right) \leq \Delta \cdot \left(\log\left(\frac{\pi_R}{1-\pi_R}\right) - \log\left(\frac{\pi_P}{1-\pi_P}\right)\right)$

Functions: *RET.Bin.LogOdds()*

RET.Bin.LogOdds.OptAlloc()

RET.Bin.LogOdds.Samplesize()

POISSON ENDPOINTS (Appendix A.2):

Model: $X_{ki} \sim Pois(\lambda_k)$ for $k = T, R, P$ and $i = 1, \ldots, n_k$.

Hypothesis: $H_{0,-\lambda_k} : \lambda_P - \lambda_T \leq \Delta \cdot (\lambda_P - \lambda_R)$

Functions: *RET.Pois.MinusId()*

RET.Pois.MinusId.OptAlloc()

RET.Pois.MinusId.Samplesize()

NORMALLY DISTRIBUTED ENDPOINTS (Appendix A.3):

Model: $X_{ki} \sim N(\mu_k, \sigma^2)$ for $k = T, R, P$ and $i = 1, \ldots, n_k$.

Hypothesis: $H_{0,\mu_k} : \mu_T - \mu_P \leq \Delta \cdot (\mu_R - \mu_P)$

Functions: *RET.Norm.Id()*

RET.Norm.Id.Samplesize()

CENSORED, EXPONENTIALLY DISTRIBUTED ENDPOINTS (Appendix A.4):

Model: (X_{ki}, δ_{ki}) with $X_{ki} = \min\{T_{ki}, U_{ki}\}$ and $\delta_{ki} = \mathbf{1}_{\{T_{ki} \leq U_{ki}\}}$ for $k = T, R, P$ and $i = 1, \ldots, n_k$, where T_{ki} is a exponentially distributed survival time with expectation $E(T_{ki}) = \lambda_k$ and U_{ki} is a censoring time, which distributions is independent of T_{ki} and not further specified.

Hypothesis: $H_{0,-\log(\lambda_k)} : \log(\lambda_P) - \log(\lambda_T) \leq \Delta \cdot (\log(\lambda_P) - \log(\lambda_R))$

Functions: *RET.Exp.MinusLog()*

RET.Exp.MinusLog.Samplesize()

DISCUSSION

In this work we have presented a full analysis and planning of three-armed trials for general retention of effect hypotheses. The endpoint of interest may follow any (regular) parametric distribution family. As a major result, we have derived the asymptotically optimal allocation, see Equation (4.1), and sample size formulas for planning the trial (4.2) and (4.3) for restricted as well as unrestricted estimation of the variance. To this end, the crucial step was the determination of the exact limit σ_{RML}^2 of the restricted MLE of the variance σ^2, which to our knowledge has not been investigated and incorporated in this context so far in the literature. As a consequence, note that while planning a non-inferiority trial it is important to decide in advance which estimation method will be performed as it affects the power and hence the total number of samples required.

For binomial endpoints our procedures improves on existing ones. This includes the precision of the sample size formula as well as the issue of optimal allocation. The optimal allocation reduces the total sample size by amounts up to 10% (20%) compared to the 2:2:1 (balanced) allocation. In addition, the methods of this work are applied to Poisson endpoints which to our knowledge have so far not been investigated in the context of three-armed non-inferiority trials, and the results from Mielke et al. (2008) for censored, exponentially distributed endpoints are summarized in the context of the general setting of this work.

A problematic issue might be that the sample size planning and evaluation of a study presented in this paper is based on asymptotically considerations. Thus, the optimal allocation can differ for finite samples, when the variance is estimated with restriction to the null hypothesis. However, this is not the case in the examples of binary endpoints with "classical" hypothesis and Poisson endpoints investigated in this work, or at least numerical studies have shown that the differences are irrelevantly small. Still, differences can occur e.g. when the ratio σ_{RML}/σ_0 is far away from 1 and the signal-to-noise ratio $\eta(0)/\sigma_0$ is very small.

Indeed, for binary endpoints with the log odds hypothesis (2.2) it turns out that, in contrast

to the retention of effect hypotheses considered so far, the asymptotically optimal allocation
of samples significantly deviates from the finite sample optimal allocation and, thus, does
not yield a satisfactory finite sample approximation, when $\pi_{0,T} = \pi_{0,R} \gg \pi_{0,P}$. However, it
figured out as long as $\pi_{0,T} = \pi_{0,R}$ is not very much larger than $\pi_{0,P}$ that the simplified optimal
allocation $n_T : n_R : n_P = 1 : \Delta : |1 - \Delta|$ is nearly optimal in the finite setting and, thus, the
recommended choice of allocation for a broad field of parameter constellations. Within the
finite sample investigations in Section 6.2.1 we determine this regions, where the simplified
optimal allocation can be applied without a significant reduction in power, and therewith also
the region, where further numerical investigations of the optimal allocation are necessary.

Beside the retention of effect hypotheses extensively discussed in this work there are evidently
further non-inferiority hypotheses of practical interest, where questions regarding sample size
planning are still open. For example, Munk et al. (2007) consider for binary endpoints the
general null hypotheses of the type

$$H^U : \pi_3 \geq h_1(\pi_1) \quad \text{or} \quad \pi_3 \geq h_2(\pi_2)$$

and

$$H^I : \pi_3 \geq h_1(\pi_1) \quad \text{and} \quad \pi_3 \geq h_2(\pi_2),$$

where π_i represents a failure or success probability under treatment i, $i = 1, 2, 3$, and h_1 and
h_2 are strictly isotonic functions. Munk et al. (2007) highlight that for specific choices of h_1
and h_2 this includes hypotheses on the differences, the relative risks or the odds ratio of the
parameters. The hypothesis H^U captures e.g. the hypothesis of demonstrating that a new
treatment is non-inferior to two standard treatments, whereas the hypothesis H^I is suitable
for showing that a new treatment is as effective as one of two given standards, or that one of
two treatments is as effective as a standard, the latter being for example the aim in Hesketh
et al. (1996). Furthermore, the extension of H^U to k-samples (Balabdaoui, Mielke and Munk,
2009), i.e. the comparison of treatment to an arbitrary number of other treatments, can be
found in dose finding studies.
The hypotheses H^U and H^I cannot be formulated in one contrast as the retention of effect
hypothesis (1.2). Therefore, union intersection and intersection union tests (see Berger, 1997),
respectively, or the likelihood ratio test, are considered to come up with a test decision. As
test procedures are developing, there are questions regarding sample size planning that remain
open, in particular the question of the optimal allocation. Here, we cannot easily derive the
limit distribution under an alternative as it is the case in the settings of this work. Note
that the likelihood ratio converges to zero under a parameter in the alternative. To this end,

Balabdaoui et al. (2009) derive the asymptotic distribution of the likelihood ratio statistic for H^U with k-samples under a local alternative. Another approach fixes the alternative and considers a higher-order expansion of the likelihood ratio statistic to obtain a non-degenerated limit distribution (see Mielke, 2007). As it is not clear (at least not clear to us) which approach should be favored or whether the results for questions of optimal sample allocation coincide, this issue could be of interest for further research.

This work essentially focuses on developing statistical methodologies and the discussion of their implementations. However, there are crucial issues related to non-inferiority trials, which are subjected to an ongoing discussion and require the expertise and the communication of statisticians, physicians and specialist from the regulatory agencies. To this end, it should be emphasized that in non-inferiority trials the specification of an appropriate non-inferiority margin (Δ) is a fundamental issue. Referring to this, Lange and Freitag (2005) claim that the selection of the non-inferiority margin has in principle the same importance for the interpretation of the statistical result as the significance level. It is not our aim or the objective of this work to take up this discussion. We want to spark the reader's interest in issues of non-inferiority trials which are beyond this work. Confronted with the animated ongoing discussion in the background we are looking forward to a first draft of the guideline for non-inferiority trials by the FDA (Food and Drug Administration, Government agency of the USA), which is expected to be soon published.

Bibliography

BALABDAOUI, F., MIELKE, M. and MUNK, A. (2009). The likelihood ratio test for non-standard hypotheses near the boundary of the null - with application to the assessment of non-inferiority. *Statistics & Decisions*, **27** 75–92.

BERGER, R. (1997). Likelihood ratio tests and intersection-union. *In: Advances in Statistical Decision Theory and Applications (S. Panchapakesan & N. Balakrishnam, eds.)* 225–237.

CHAN, I. (1998). Exact tests of equivalence and efficacy with a non-zero lower bound for comparative studies. *Statistics in Medicine*, **17** 1403–1413.

D'AGOSTINO, R. (2003). Special issue: Non-inferiority trials: Advances in concepts and methodology. *Statistics in Medicine*, **22** 165–336.

DUNNETT, C. W. and GENT, M. (1977). Significance testing to establish equivalence between treatments, with special reference to data in the form of 2x2 tables. *Biometrics*, **33** 593–602.

DWORKIN, R., KATZ, J. and GITLIN, M. (2005). Placebo response in clinical trials of depression and its implications for research on chronic neuropathic pain. *Neurology*, **65** 7–19.

FARRINGTON, C. and MANNING, G. (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine*, **9** 1447–1454.

GOLDSTEIN, D., LU, Y., DETKE, M., WILTSE, C., MALLINCKRODT, C. and DEMITRACK, M. (2004). Duloxetine in the treatment of depression. a double-blind placebo-controlled comparison with paroxetine. *Journal of Clinical Psychopharmacology*, **24** 389–398.

GURM, H. S., YADAV, J. S., FAYAD, P., KATZEN, B. T., MISHKEL, G. J., BAJWA, T. K., ANSEL, G., STRICKMAN, N. E., WANG, H., COHEN, S. A., MASSARO, J. M. and CUTLIP, D. E. (2008). Long-term results of carotid stenting versus endarterectomy in high-risk patients. *New England Journal of Medicine*, **358** 1572–1579.

HASLER, M., VONK, R. and HOTHORN, L. (2008). Assessing non-inferiority of a new treatment in a three-arm trial in the presence of heteroscedasticity. *Statistics in Medicine*, **27** 490–503.

HASSELBLAD, V. and KONG, D. (2001). Statistical methods for comparison to placebo in active-control trials. *Drug Information Journal*, **35** 435–449.

HAUCK, W. and ANDERSON, S. (1999). Some issues in the design and analysis of equivalence trials. *Drug Information Journal*, **33** 109–118.

HAUSCHKE, D. and PIGEOT, I. (2005). Establishing efficacy of a new experimental treatment in the "gold standard" design. *Biometrical Journal*, **47** 782–786.

HESKETH, P., NAVARI, R., GROTE, T., GRALLA, R., HAINSWORTH, J., KRIS, M., ANTHONY, L., KHOJASTEH, A., TAPAZOGLOU, E., C., B. and HAHNE, W. (1996). Double-blind, randomized comparison of the antiemetic efficacy of intravenous dolasetron mesylate and intravenous ondansetron in the prevention of acute cisplatin-induced emesis in patients with cancer. dolasetron comparative chemotherapy-induced emesis prevention group. *Journal of Clinical Oncology*, **14** 2242–2249.

HOLGREM, E. (1999). Establishing equivalence by showing that a specified percentage of the effect of the active control over placebo is maintained. *Journal of Biopharmaceutical Statistics*, **9** 651–659.

HUNG, H., WANG, S. and O'NEILL, R. (2009). Challenges and regulatory experiences with noninferiority trial design without placebo arm. *Biometrical Journal*, **51** 324–334.

HUNG, J., WANG, S.-J. and O'NEIL, R. (2005). A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials. *Biometrical Journal*, **47** 28–36.

HYPERICUM DEPRESSION TRIAL STUDY GROUP (2004). Effect of hypericum perforatum (st john's wort) in major depressive disorder. *Journal of the American Medical Association*, **287** 1807–1814.

ICH (2000). ICH Harmonized Tripartite Guideline. Choice of Control Group and Related Issues in Clinical Trials (E10). *CPMP/ICH/364/96*.

JENNRICH, R. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, **40** 633–643.

JONES, B., JARVIS, P., LEWIS, J. and EBBUTT, A. (1996). Trials to assess equivalence: the importance of rigorous methods. *British Medical Journal*, **313** 36–39.

Kieser, M. and Friede, T. (2007). Planning and analysis of three-arm non-inferiority trial with binary endpoints. *Statistics in Medicine*, **26** 253–273.

Kim, M. and Xue, X. (2004). Likelihood ratio and a bayesian approach were superior to standard noninferiority analysis when the noninferiority margin varied with the control event rate. *Journal of Clinical Epidemiology*, **57** 1253–1261.

Koch, A. (2005). Discussion on "establishing efficacy of a new experimental treatment in the 'gold standard' design". *Biometrical Journal*, **47** 792–793.

Koch, A. and Röhmel, J. (2004). Hypothesis testing in the "gold standard" design for proving the efficacy of an experimental treatment relative to placebo and a reference. *Journal of Biopharmaceutical Statistics*, **14** 315–325.

Koch, G. and Tangen, C. (1999). Nonparametric analysis of covariance and its role in noninferiority clinical trials. *Drug Information Journal*, **33** 1145–1159.

Lange, S. and Freitag, G. (2005). Choice of delta: Requirements and reality - results of a systematic review. *Biometrical Journal*, **47** 12–27.

Layard, M. and Arvesen, J. (1978). Analysis of poisson data in crossover experimental designs. *Biometrics*, **34** 421–428.

Lui, K.-J. (2005). Sample size calculation for testing non-inferiority and equivalence under poisson distribution. *Statistical Methodology*, **2** 37–48.

Mielke, M. (2007). *Die asymptotische Verteilung des Likelihood-Quotienten-Tests für allgemeine Hypothesenräume*. Diploma thesis, Inst. for Math. Stochastics, Georg-August-University of Göttingen, Göttingen.

Mielke, M. and Munk, A. (2010). The assessment and planning of non-inferiority trials for retention of effect hypotheses - towards a general approach. *Submitted*.

Mielke, M., Munk, A. and Schacht, A. (2008). Planning and assessing non-inferiority in a gold standard design with censored, exponentially distributed endpoints. *Statistics in Medicine*, **27** 5093–5110.

Mohanraj, R. and Brodie, M. (2003). Measuring the efficacy of antiepileptic drugs. *Seizure*, **12** 413–443.

Munk, A., Mielke, M., Skipka, G. and Freitag, G. (2007). Testing noninferiority in three-armed clinical trials based on the likelihood ratio statistics. *Canadian Journal of Statistics*, **35** 413–431.

MUNK, A. and TRAMPISCH, H.-J. (2005). Special issue: Therapeutic equivalence - clinical issues and statistical methodology in noninferiority trials. *Biometrical Journal*, **47** 1–108.

NIERENBERG, A. and WRIGHT, E. (1999). Evolution of remission as the new standard in the treatment of depression. *Journal of Clinical Psychiatry*, **60 (suppl. 22)** 7–11.

PIGEOT, I., SCHÄFER, J., RÖHMEL, J. and HAUSCHKE, D. (2003). Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine*, **22** 883–899.

RÖHMEL, J. (1998). Therapeutic equivalence investigations: statistical considerations. *Statistics in Medicine*, **17** 1703–1714.

RÖHMEL, J. and MANSMANN, U. (1999). Unconditional non-asymptotic one-sided test for independent binomial proportions when the interest lies in showing non-inferiority and/or superiority. *Biometrical Journal*, **41** 149–170.

ROTHMANN, M., LI, N., CHEN, G., CHI, G., TEMPLE, R. and TSOU, H.-H. (2003). Design and analysis of non-inferiority mortality trials in oncology. *Statistics in Medicine*, **22** 239–264.

SANDER, J., PATSALOS, P., OXLEY, J., HAMILTON, M. and YUEN, W. (1990). A randomized double-blind placebo-controlled add-on trial of lamotrigine in patients with severe epilepsy. *Epilepsy Res.*, **6** 221–226.

SCHWARTZ, T. and DENNE, J. (2006). A two-stage sample size recalculation procedure for placebo- and active-controlled non-inferiority trials. *Statistics in Medicine*, **45** 3396–3406.

SENN, S. (1997). *Statistical Issues in Drug Development*. John Wiley & Sons.

SKIPKA, G., MUNK, A. and FREITAG, G. (2004). Unconditional exact tests for the difference of binomial probabilities - contrasted and compared. *Computational Statistics & Data Analysis*, **47** 757–773.

TANG, N.-S., TANG, M.-L. and WANG, S.-F. (2007). Sample size determination for matched-pair equivalence trials using rate ratio. *Biostatistics*, **8** 625–631.

TEMPLE, R. and ELLENBERG, S. (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. part 1: Ethical and scientific issues. *Annals of Internal Medicine*, **133** 455–463.

VAN DER VAART, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

WALD, A. (1949). Note on consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, **20** 595–601.

WALSH, B., SEIDMAN, S. and SYSKO, R. (2002). Placebo response in studies of major depression - variable, substantial and growing. *Journal of the American Medical Association*, **287** 1840–1847.

WHITE, H. (1980). Nonlinear regression on cross-section data. *Econometrica*, **48** 721–746.

WMA (2000). World medical association declaration of helsinki - ethical principles for medical research involving human subjects. *http://www.wma.net/e/policy/b3.htm*.

YADID, G., ZANGEN, A., DMITROCHENKO, A., OVERSTREET, D. and ZOHAR, J. (2000). Screening for new antidepressants with fast onset and long-lasting action. *Drug Development Research*, **50** 392–399.

YAKHNO, N., GUEKHT, A., SKOROMETS, A., SPIRIN, N., STRACHUNSKAYA, E., TERNAVSKY, A., OLSEN, K. and MOLLER, P. (2006). Lornoxicam quick-release compared with diclofenac potassium: subjects and methods. *Clinical Drug Investigation*, **26** 267–277.

ZHANG, Z. (2006). Non-inferiority testing with a variable margin. *Biometrical Journal*, **48** 948–965.

IMPLEMENTATIONS IN R

A.1   Binary endpoints

A.1.1   "Classical" retention of effect hypothesis

```
1  ###############################################################################
   #————————————————————————————————————————————————————————————————————————#
3  #                                                                          #
   # BINARY ENDPOINTS                                                         #
5  #                                                                          #
   ###############################################################################
7
   #==========================================================================
9  #   FUNCTION: RET.Bin.Id
   # ————————————————————————————————————————————————————————————————————————
11 #   INPUT: x.t           Successes in the test group
   #          x.r           Successes in the reference group
13 #          x.p           Successes in the placebo group
   #          n.t           Total number of samples in the test group
15 #          n.r           Total number of samples in the reference group
   #          n.p           Total number of samples in the placebo group
17 #          delta         Non-inferiority margin:
   #                        Fraction of active controll effect which should be
19 #                        retained by the test treatment relative to placebo
   #          proc          =c("restricted","unrestricted")
21 #                        "restricted": the variance sigma^2 is estimated
   #                             by restricted ML to the null-hypothesis
23 #                        "unrestricted": the variance sigma^2 is estimated
   #                             by unrestricted ML
25 # ————————————————————————————————————————————————————————————————————————
   #   OUTPUT: proc          Variance estimating procedure
27 #           teststatistic Value of the teststatistic T
   #           p.value       To the teststatistic corresponding p-value
29 #
   #==========================================================================
31 RET.Bin.Id <- function(x.t,x.r,x.p,
                          n.t,n.r,n.p,
33                        delta,
                          proc=c("restricted","un.restricted"))
35 {
   # MLE's ————————————————————————————————————————————————————————————————
37  pi.ml.t <- x.t/n.t
```

```r
     pi.ml.r <- x.r/n.r
39   pi.ml.p <- x.p/n.p

41   # Estimation of the variance sigma^2 ————————————————————————————
     if (proc=="restricted"){
43       # Computation of the restricted MLE's ——————————————————————

45       # likelihood in one group ——————————————————————————————
             l.one <- function(p,x,n){
47             res <- choose(n,x)*(p^x)*((1-p)^(n-x))
               return(res)
49           }
         # common likelihood ———————————————————————————————————
51           ln <- function(p.t,p.r,p.p,x.t,x.r,x.p,n.t,n.r,n.p){
               res <- (l.one(p.t,x.t,n.t)) *(l.one(p.r,x.r,n.r))*(l.one(p.p,x.p,n.p))
53             return(res)
             }
55       # minus restricted likelihood ————————————————————————————
             ln.res <- function(par,x.t,x.r,x.p,n.t,n.r,n.p,d){
57             pi.t <- d*par[1]+(1-d)*par[2]
               res <- -ln(pi.t,par[1],par[2],x.t,x.r,x.p,n.t,n.r,n.p)
59             return(res)
             }
61

         res <- optim(c(pi.ml.r,pi.ml.p),ln.res,x.t=x.t,x.r=x.r,x.p=x.p,n.t=n.t,
63           n.r=n.r,n.p=n.p,d=delta,method = "L-BFGS-B",lower=c(0,0),upper=c(1,1))

65       pi.rml.r <- res$par[1]
         pi.rml.p <- res$par[2]
67       pi.rml.t <- delta*pi.rml.r + (1-delta)*pi.rml.p

69       sigma <- sqrt(pi.rml.t*(1-pi.rml.t)/n.t + delta^2*pi.rml.r*(1-pi.rml.r)/n.r+
                         (1-delta)^2*pi.rml.p*(1-pi.rml.p)/n.p)
71   }
     if (proc=="unrestricted"){
73     sigma <- sqrt(pi.ml.t*(1-pi.ml.t)/n.t + delta^2*pi.ml.r*(1-pi.ml.r)/n.r+
                       (1-delta)^2*pi.ml.p*(1-pi.ml.p)/n.p)
75   }
     # Teststatistic ————————————————————————————————————————————
77   teststatistic <- (pi.ml.t - delta*pi.ml.r-(1-delta)*pi.ml.p)/sigma
     #————————————————————————————————————————————————————————————
79   result <- cbind(proc,round(teststatistic,4),round(1-pnorm(teststatistic),4))
     colnames(result) <- c("proc","teststatistic","p.value")
81   result <- as.data.frame(result)
     return(result)
83 }


85 #————————————————————————————————————————————————————————————————
   #   EXAMPLE:
87 #   Performing the RET for an example in the treatment of depression:
   #
89 #       Treatment      Successes      No. of patients
   #   ————————————————————————————————————————————
91 #        T(est)           43                 86
   #      R(eference)        31                 84
93 #       P(lacebo)         26                 88
   #
95 #     (see Section 2.1, Table 2.1)
   #————————————————————————————————————————————————————————————————
97 RET.Bin.Id(x.t=43,x.r=31,x.p=26,n.t=86,n.r=84,n.p=88,delta=0.8,
```

```
                                                          proc="restricted")
 99 RET.Bin.Id(x.t=43,x.r=31,x.p=26,n.t=86,n.r=84,n.p=88,delta=0.8,
                                                          proc="unrestricted")

101
    #————————————————————————————————————————————————————————————————————
103 # OUTPUT:
    #
105 #> RET.Bin.Id(x.t=43,x.r=31,x.p=26,n.t=86,n.r=84,n.p=88,delta=0.8,
    # +                                        proc="restricted")
107 #        proc      T p.value
    # 1 restricted 2.1034  0.0177
109 #
    #> RET.Bin.Id(x.t=43,x.r=31,x.p=26,n.t=86,n.r=84,n.p=88,delta=0.8,
111 # +                                        proc="unrestricted")
    #          proc       T p.value
113 # 1 unrestricted 2.1079  0.0175
    #————————————————————————————————————————————————————————————————————
115 ##################################################################################

117 #————————————————————————————————————————————————————————————————————
    #   FUNCTION: RET.Bin.Id.OptAlloc
119 # —————————————————————————————————————————————————————————————————————
    #   INPUT:    pi.t      Success probability in the test group
121 #             pi.r      Success probability in the reference group
    #             pi.p      Success probability in the placebo group
123 #             delta     Non-inferiority margin:
    #                       Fraction of active controll effect which should be
125 #                       retained by the test treatment relative to placebo
    # —————————————————————————————————————————————————————————————————————
127 #   OUTPUT:          Optimal sample allocation:
    #             w.t      Relative proportion of the test group        |
129 #             w.r      Relative proportion of the reference group    | rounded on
    #             w.p      Relative proportion of the placebo group      | 3 digits
131 #
    #————————————————————————————————————————————————————————————————————
133 RET.Bin.Id.OptAlloc <- function(pi.t,pi.r,pi.p,delta){

135   w.t <- sqrt(pi.t*(1-pi.t))
      w.r <- delta*sqrt(pi.r*(1-pi.r))
137   w.p <- abs(1-delta)*sqrt(pi.p*(1-pi.p))

139   result <- round(cbind(w.t,w.r,w.p)/(w.t+w.r+w.p),3)
      colnames(result) <- c("w.t","w.r","w.p")
141   result <- as.data.frame(result)

143   return(result)
    }
145
    #————————————————————————————————————————————————————————————————————
147 #   EXAMPLE
    #
149 #   Optimal allocation for the RET under the alternative
    #   pi.t=pi.r=0.5, pi.p=0.1 and delta=70%
151 #————————————————————————————————————————————————————————————————————
    RET.Bin.Id.OptAlloc(pi.t=0.5,pi.r=0.5,pi.p=0.1,delta=0.7)
153
    #————————————————————————————————————————————————————————————————————
155 # OUTPUT:
    # RET.Bin.Id.OptAlloc(pi.t=0.5,pi.r=0.5,pi.p=0.1,delta=0.7)
157 #        w.t    w.r    w.p
```

```
      #  1 0.532 0.372 0.096
159 #—————————————————————————————————————————————————————
      ##################################################################################
161
      #===============================================================================
163 #   FUNCTION: RET. Bin. Id. Samplesize
      # ——————————————————————————————————————————————————————
165 #   INPUT:     pi.t       Success probability in the test group
      #             pi.r       Success probability in the reference group
167 #             pi.p       Success probability in the placebo group
      #             w.t        Relative sample allocation to the test group (n.t/n)
169 #             w.r        Relative sample allocation to the reference group (n.r/n)
      #             delta      Non-inferiority margin:
171 #                        Fraction of active controll effect which should be
      #                        retained by the test treatment relative to placebo
173 #             level      Significance level
      #             pow        Aspired power (one minus the type II-error)
175 #             proc       =c("restricted","unrestricted")
      #                        "restricted": the variance sigma^2 will be estimated
177 #                             by restricted ML to the null-hypothesis
      #                        "unrestricted": the variance sigma^2 will be estimated
179 #                             by unrestricted ML
      # ——————————————————————————————————————————————————————
181 #   OUTPUT:    proc       Variance estimating procedure, which will be used
      #
183 #                        Required sample sizes:
      #             n          Total required sample size (n.t+n.r+n.p)
185 #             n.t        Required samples for the test group (rounded)
      #             n.r        Required samples for the refernce group (rounded)
187 #             n.p        Required samples for the placebo group (rounded)
      #
189 #===============================================================================
    RET. Bin. Id. Samplesize <- function(pi.t, pi.r, pi.p,
191                                        w.t, w.r,
                                           delta, level, pow,
193                                        proc=c("restricted","unrestricted"))
    {
195   if (proc=="restricted"){
      #—— Kullback-Leibler divergence ——————————————————————————
197     KL <- function(pi.0, p)
        {
199       pi.0*log(pi.0/p)+(1-pi.0)*log((1-pi.0)/(1-p))
        }
201   #—— weighted Kullback-Leibler divergence ——————————————————
        weighted.KL <- function(par, pi.t.0, pi.r.0, pi.p.0, w.t, w.r, d)
203     {
          pi.r <- par[1]
205       pi.p <- par[2]

207       w.KL <- w.t*KL(pi.t.0, d*pi.r+(1-d)*pi.p) +
                  w.r*KL(pi.r.0, pi.r) + (1-w.r-w.t)*KL(pi.p.0, pi.p)
209       return(w.KL)
        }
211   #————————————————————————————————————————————————————————
        # KL-minimizer ————————————————————————————————————————
213     par <- optim(c(pi.r, pi.p), weighted.KL,
                     pi.t.0=pi.t, pi.r.0=pi.r, pi.p.0=pi.p,
215                  w.t=w.t, w.r=w.r, d=delta)$par
        pi.rml.r <- par[1]
217     pi.rml.p <- par[2]
```

```
          pi.rml.t <- delta*pi.rml.r + (1-delta)*pi.rml.p
219
          sigma.sq.rml <-   pi.rml.t*(1-pi.rml.t)/w.t +
221                          (delta^2)*pi.rml.r*(1-pi.rml.r)/w.r +
                            (1-delta)^2 * pi.rml.p*(1-pi.rml.p)/(1-w.r-w.t)
223
          sigma.sq <-   pi.t*(1-pi.t)/w.t +
225                       (delta^2)*pi.r*(1-pi.r)/w.r +
                         (1-delta)^2 * pi.p*(1-pi.p)/(1-w.r-w.t)
227
          eta <- pi.t-delta*pi.r-(1-delta)*pi.p
229
          z.alpha <- qnorm(level)
231       z.beta  <- qnorm(1-pow)

233       n <- (z.alpha*sqrt(sigma.sq.rml)+z.beta*sqrt(sigma.sq))^2/eta^2
      }
235   if (proc=="unrestricted"){

237       sigma.sq <-   pi.t*(1-pi.t)/w.t +
                         (delta^2)*pi.r*(1-pi.r)/w.r +
239                      (1-delta)^2*pi.p*(1-pi.p)/(1-w.r-w.t)

241       eta <- pi.t-delta*pi.r-(1-delta)*pi.p

243       z.alpha <- qnorm(level)
          z.beta  <- qnorm(1-pow)
245
          n <- (z.alpha*sqrt(sigma.sq)+z.beta*sqrt(sigma.sq))^2/eta^2
247   }
      # rounding the sample sizes in each group ─────────────────────
249   n.t <- round(n*w.t)
      n.r <- round(n*w.r)
251   n.p <- round(n*(1-w.t-w.r))

253   result <- cbind(proc,n.t+n.r+n.p,n.t,n.r,n.p)

255   colnames(result) <- c("proc","n","n.t","n.r","n.p")
      result <- as.data.frame(result)
257
      return(result)
259 }

261 #═══════════════════════════════════════════════════════════════
    #   EXAMPLE
263 #
    #   Determine the minimal required sample size for the alternative
265 #   pi.t=pi.r=0.5, pi.p=0.1 and delta=70% under the optimal sample
    #   allocation (w.t,w.r,w.p)=(0.532,0.372,0.096) computed with
267 #   'RET.Bin.Id.OptAlloc()' when the variance sigma^2 will be estimated
    #   with restriction to null hypothesis.
269 #═══════════════════════════════════════════════════════════════
    RET.Bin.Id.Samplesize(pi.t=0.5,pi.r=0.5,pi.p=0.1,w.t=0.532,w.r=0.372,
271                                    delta=0.7,level=0.05,pow=0.8,
                                      proc="restricted")
273
    #───────────────────────────────────────────────────────────────
275 # OUTPUT:
    #>   RET.Bin.Id.Samplesize(pi.t=0.5,pi.r=0.5,pi.p=0.1,w.t=0.532,w.r=0.372,
277 # +                                delta=0.7,level=0.05,pow=0.8,
```

```
   # +                                           proc="restricted")
279 #          proc    n n.t n.r n.p
   # 1 restricted 387 206 144   37
281 #
   ################################################################################
```

A.1.2 Log odds retention of effect hypothesis

```
 1 #
   #   FUNCTION: RET.Bin.LogOdds
 3 #
   #   INPUT: x.t          Successes in the test group
 5 #          x.r          Successes in the reference group
   #          x.p          Successes in the placebo group
 7 #          n.t          Total number of samples in the test group
   #          n.r          Total number of samples in the reference group
 9 #          n.p          Total number of samples in the placebo group
   #          delta        Non-inferiority margin:
11 #                       Fraction of active controll effect which should be
   #                       retained by the test treatment relative to placebo
13 #          proc         =c("restricted","unrestricted")
   #                       "restricted": the variance sigma^2 is estimated
15 #                           by restricted ML to the null-hypothesis
   #                       "unrestricted": the variance sigma^2 is estimated
17 #                           by unrestricted ML
   #
19 #   OUTPUT: proc          Variance estimating procedure
   #           teststatistic  Value of the teststatistic T
21 #           p.value        To the teststatistic corresponding p-value
   #
23 #
   RET.Bin.LogOdds <- function(x.t,x.r,x.p,
25                              n.t,n.r,n.p,
                               delta,
27                              proc=c("restricted","unrestricted"))
   {
29  # MLE's
   pi.ml.t <- x.t/n.t
31  pi.ml.r <- x.r/n.r
   pi.ml.p <- x.p/n.p
33
   h <- function(pi){
35   value <- log(pi/(1-pi))
    return(value)
37  }

39  pi.t.b <- function(pi.r,pi.p,d){
     res <- exp(d*h(pi.r)+(1-d)*h(pi.p))
41    value <- res/(1+res)
     return(value)
43  }

45  # Estimation of the variance sigma^2
   if (proc=="restricted"){
47    # Computation of the restricted MLE's

49      # likelihood in one group
         l.one <- function(p,x,n){
51          res <- choose(n,x)*(p^x)*((1-p)^(n-x))
            return(res)
```

```
53          }
        # common likelihood ──────────────────────────────────────
55        ln <- function(p.t,p.r,p.p,x.t,x.r,x.p,n.t,n.r,n.p){
             res <- (l.one(p.t,x.t,n.t)) *(l.one(p.r,x.r,n.r))*(l.one(p.p,x.p,n.p))
57           return(res)
          }
59        # minus restricted likelihood ─────────────────────────────
          ln.res <- function(par,x.t,x.r,x.p,n.t,n.r,n.p,d){
61           pi.t <- pi.t.b(par[1],par[2],d)
             res <- -ln(pi.t,par[1],par[2],x.t,x.r,x.p,n.t,n.r,n.p)
63           return(res)
          }
65
        res <- optim(c(pi.ml.r,pi.ml.p),ln.res,x.t=x.t,x.r=x.r,x.p=x.p,n.t=n.t,
67          n.r=n.r,n.p=n.p,d=delta,method = "L-BFGS-B",lower=c(0,0),upper=c(1,1))

69      pi.rml.r <- res$par[1]
        pi.rml.p <- res$par[2]
71      pi.rml.t <- pi.t.b(pi.rml.r,pi.rml.p,delta)

73      sigma <- sqrt(1/(pi.rml.t*(1-pi.rml.t)*n.t) + delta^2/
                        (pi.rml.r*(1-pi.rml.r)*n.r)+
75                      (1-delta)^2/(pi.rml.p*(1-pi.rml.p)*n.p))
      }
77   if (proc=="unrestricted"){
        sigma <- sqrt(1/(pi.ml.t*(1-pi.ml.t)*n.t) + delta^2/
79                      (pi.ml.r*(1-pi.ml.r)*n.r)+
                        (1-delta)^2/(pi.ml.p*(1-pi.ml.p)*n.p))
81   }
     # Teststatistic ──────────────────────────────────────────────
83   teststatistic <- (h(pi.ml.t) - delta*h(pi.ml.r)-(1-delta)*h(pi.ml.p))/sigma
     #──────────────────────────────────────────────────────────────
85   result <- cbind(proc,round(teststatistic,4),round(1-pnorm(teststatistic),4))
     colnames(result) <- c("proc","teststatistic","p.value")
87   result <- as.data.frame(result)
     return(result)
89 }


91 #══════════════════════════════════════════════════════════════════
   #   EXAMPLE:
93 #   Performing the RET for an example in the treatment of depression:
   #
95 #      Treatment     Successes      No. of patients
   #    ─────────────────────────────────────────────────
97 #       T(est)         43               86
   #     R(eference)      31               84
99 #      P(lacebo)       26               88
   #
101 #    (see Section 2.1, Table 2.1)
   #══════════════════════════════════════════════════════════════════
103 RET.Bin.LogOdds(x.t=43,x.r=31,x.p=26,n.t=86,n.r=84,n.p=88,delta=0.8,
                                        proc="restricted")
105 RET.Bin.LogOdds(x.t=43,x.r=31,x.p=26,n.t=86,n.r=84,n.p=88,delta=0.8,
                                        proc="unrestricted")

107
   #──────────────────────────────────────────────────────────────────
109 # OUTPUT:
   #
111 #> RET.Bin.LogOdds(x.t=43,x.r=31,x.p=26,n.t=86,n.r=84,n.p=88,delta=0.8,
   #+                                       proc="restricted")
```

```
113 #          proc  teststatistic  p.value
    #1  restricted          2.1183    0.0171
115 #> RET.Bin.LogOdds(x.t=43,x.r=31,x.p=26,n.t=86,n.r=84,n.p=88,delta=0.8,
    #+                                     proc="unrestricted")
117 #          proc  teststatistic  p.value
    #1  unrestricted          2.1128    0.0173
119 #—————————————————————————————————————————————————————————————
    #################################################################################
121
    #═══════════════════════════════════════════════════════════════
123 #   FUNCTION: RET.Bin.LogOdds.OptAlloc
    # ——————————————————————————————————————————————————————————————
125 #   INPUT:     pi.t        Success  probability  in  the  test  group
    #              pi.r        Success  probability  in  the  reference  group
127 #              pi.p        Success  probability  in  the  placebo  group
    #              delta       Non—inferiority  margin:
129 #                          Fraction  of  active  controll  effect  which  should  be
    #                          retained  by  the  test  treatment  relative  to  placebo
131 # ——————————————————————————————————————————————————————————————
    #   OUTPUT:             Optimal  sample  allocation:
133 #              w.t        Relative  proportion  of  the  test  group        |
    #              w.r        Relative  proportion  of  the  reference  group   | rounded on
135 #              w.p        Relative  proportion  of  the  placebo  group      | 3 digits
    #
137 #═══════════════════════════════════════════════════════════════
    RET.Bin.LogOdds.OptAlloc <- function(pi.t,pi.r,pi.p,delta){
139
      w.t <- 1/sqrt(pi.t*(1-pi.t))
141   w.r <- delta/sqrt(pi.r*(1-pi.r))
      w.p <- abs(1-delta)/sqrt(pi.p*(1-pi.p))
143
      result <- round(cbind(w.t,w.r,w.p)/(w.t+w.r+w.p),3)
145   colnames(result) <- c("w.t","w.r","w.p")
      result <- as.data.frame(result)
147
      return(result)
149 }

151 #═══════════════════════════════════════════════════════════════
    #   EXAMPLE
153 #
    #   Optimal  allocation  for  the  RET  under  the  alternative
155 #   pi.t=pi.r=0.5,  pi.p=0.1  and  delta=70%
    #═══════════════════════════════════════════════════════════════
157 RET.Bin.LogOdds.OptAlloc(pi.t=0.5,pi.r=0.5,pi.p=0.1,delta=0.7)

159 #——————————————————————————————————————————————————————————————
    # OUTPUT:
161 #> RET.Bin.LogOdds.OptAlloc(pi.t=0.5,pi.r=0.5,pi.p=0.1,delta=0.7)
    #     w.t    w.r    w.p
163 #1  0.455  0.318  0.227
    #——————————————————————————————————————————————————————————————
165 #################################################################################

167 #═══════════════════════════════════════════════════════════════
    #   FUNCTION: RET.Bin.LogOdds.Samplesize
169 # ——————————————————————————————————————————————————————————————
    #   INPUT:     pi.t        Success  probability  in  the  test  group
171 #              pi.r        Success  probability  in  the  reference  group
    #              pi.p        Success  probability  in  the  placebo  group
```

```
173 #              w.t        Relative  sample  allocation  to  the  test  group  (n.t/n)
    #              w.r        Relative  sample  allocation  to  the  reference  group  (n.r/n)
175 #              delta      Non-inferiority  margin:
    #                         Fraction  of  active  controll  effect  which  should  be
177 #                         retained  by  the  test  treatment  relative  to  placebo
    #              level      Significance  level
179 #              pow        Aspired  power  (one  minus  the  type  II-error)
    #              proc       =c("restricted","unrestricted")
181 #                         "restricted":  the  variance  sigma^2  will  be  estimated
    #                              by  restricted  ML  to  the  null-hypothesis
183 #                         "unrestricted":  the  variance  sigma^2  will  be  estimated
    #                              by  unrestricted  ML
185 # ————————————————————————————————————————————————————————————————
    #  OUTPUT:  proc     Variance  estimating  procedure,  which  will  be  used
187 #
    #                     Required  sample  sizes:
189 #              n        Total  required  sample  size  (n.t+n.r+n.p)
    #              n.t      Required  samples  for  the  test  group  (rounded)
191 #              n.r      Required  samples  for  the  refernce  group  (rounded)
    #              n.p      Required  samples  for  the  placebo  group  (rounded)
193 #
    #================================================================
195 RET.Bin.LogOdds.Samplesize <- function(pi.t,pi.r,pi.p,
                                           w.t,w.r,
197                                        delta,level,pow,
                                           proc=c("restricted","unrestricted"))
199 {
    h <- function(pi){
201   value <- log(pi/(1-pi))
      return(value)
203  }

205  pi.t.b <- function(pi.r,pi.p,d){
       res <- exp(d*h(pi.r)+(1-d)*h(pi.p))
207    value <- res/(1+res)
       return(value)
209  }


211  if (proc=="restricted"){
    #—— Kullback-Leibler divergence ——————————————————————————
213    KL <- function(pi.0, p)
       {
215      pi.0*log(pi.0/p)+(1-pi.0)*log((1-pi.0)/(1-p))
       }
217  #—— weighted Kullback-Leibler divergence ——————————————————
       weighted.KL <- function(par,pi.t.0,pi.r.0,pi.p.0,w.t,w.r,d)
219    {
         pi.r <- par[1]
221      pi.p <- par[2]

223      w.KL <- w.t*KL(pi.t.0,pi.t.b(pi.r,pi.p,d)) +
                 w.r*KL(pi.r.0,pi.r) + (1-w.r-w.t)*KL(pi.p.0,pi.p)
225      return(w.KL)
       }
227  #————————————————————————————————————————————————————————
       # KL-minimizer ——————————————————————————————————————————
229    par <- optim(c(pi.r,pi.p), weighted.KL,
                    pi.t.0=pi.t,pi.r.0=pi.r,pi.p.0=pi.p,
231                 w.t=w.t,w.r=w.r,d=delta)$par
       pi.rml.r <- par[1]
```

```
233        pi.rml.p <- par[2]
           pi.rml.t <- pi.t.b(pi.rml.r,pi.rml.p,delta)

235
           sigma.sq.rml <-   1/(pi.rml.t*(1-pi.rml.t)*w.t) +
237                          (delta^2)/(pi.rml.r*(1-pi.rml.r)*w.r) +
                             (1-delta)^2 / (pi.rml.p*(1-pi.rml.p)*(1-w.r-w.t))

239
           sigma.sq <-   1/(pi.t*(1-pi.t)*w.t) +
241                      (delta^2)/(pi.r*(1-pi.r)*w.r) +
                         (1-delta)^2 / (pi.p*(1-pi.p)*(1-w.r-w.t))

243
           eta <- h(pi.t)-delta*h(pi.r)-(1-delta)*h(pi.p)

245
           z.alpha <- qnorm(level)
247        z.beta  <- qnorm(1-pow)

249        n <- (z.alpha*sqrt(sigma.sq.rml)+z.beta*sqrt(sigma.sq))^2/eta^2
      }
251   if (proc=="unrestricted"){

253        sigma.sq <-   1/(pi.t*(1-pi.t)*w.t) +
                         (delta^2)/(pi.r*(1-pi.r)*w.r) +
255                      (1-delta)^2/(pi.p*(1-pi.p)*(1-w.r-w.t))

257        eta <- h(pi.t)-delta*h(pi.r)-(1-delta)*h(pi.p)

259        z.alpha <- qnorm(level)
           z.beta  <- qnorm(1-pow)

261
           n <- (z.alpha*sqrt(sigma.sq)+z.beta*sqrt(sigma.sq))^2/eta^2
263   }
      # rounding the sample sizes in each group ——————————————
265   n.t <- round(n*w.t)
      n.r <- round(n*w.r)
267   n.p <- round(n*(1-w.t-w.r))

269   result <- cbind(proc,n.t+n.r+n.p,n.t,n.r,n.p)

271   colnames(result) <- c("proc","n","n.t","n.r","n.p")
      result <- as.data.frame(result)

273
      return(result)
275 }


277 #=====================================================================
    #   EXAMPLE
279 #
    #   Determine the minimal required sample size for the alternative
281 #   pi.t=pi.r=0.5, pi.p=0.1 and delta=70% under the optimal sample
    #   allocation (w.t,w.r,w.p)=(0.455,0.318,0.227) computed with
283 #   'RET.Bin.LogOdd.OptAllocs()' when the variance sigma^2 will be estimated
    #   with restriction to null hypothesis.
285 #=====================================================================
    RET.Bin.LogOdds.Samplesize(pi.t=0.5,pi.r=0.5,pi.p=0.2,w.t=0.5,w.r=0.3,
287                                delta=0.7,level=0.05,pow=0.8,
                                   proc="restricted")
289 RET.Bin.LogOdds.Samplesize(pi.t=0.5,pi.r=0.5,pi.p=0.2,w.t=0.5,w.r=0.3,
                                   delta=0.7,level=0.05,pow=0.8,
291                                proc="unrestricted")
```

```
293  #————————————————————————————————————————————————————————————————————————————————————
     # OUTPUT:
295  #> RET.Bin.LogOdds.Samplesize(pi.t=0.5,pi.r=0.5,pi.p=0.2,w.t=0.5,w.r=0.3,
     #+                                      delta=0.7,level=0.05,pow=0.8,
297  #+                                      proc="restricted")
     #          proc     n n.t n.r n.p
299  #1  restricted 616 308 185 123
     #> RET.Bin.LogOdds.Samplesize(pi.t=0.5,pi.r=0.5,pi.p=0.2,w.t=0.5,w.r=0.3,
301  #+                                      delta=0.7,level=0.05,pow=0.8,
     #+                                      proc="unrestricted")
303  #          proc     n n.t n.r n.p
     #1  unrestricted 620 310 186 124
305  #————————————————————————————————————————————————————————————————————————————————————
     #####################################################################################
```

A.2   Poisson endpoints

```
###############################################################################
#-----------------------------------------------------------------------------#
#                                                                             #
# POISSON ENDPOINTS                                                           #
#                                                                             #
###############################################################################

#=============================================================================
#   FUNCTION: RET.Pois.MinusId
#-----------------------------------------------------------------------------
#   INPUT:  x.t          Sum of observation or a vector with each single
#                        observation in the test group
#           x.r          Sum of observation or a vector with each single
#                        observation in the reference group
#           x.p          Sum of observation or a vector with each single
#                        observation in the placebo group
#           n.t          Total number of samples in the test group
#           n.r          Total number of samples in the reference group
#           n.p          Total number of samples in the placebo group
#           delta        Non-inferiority margin:
#                        Fraction of active controll effect which should be
#                        retained by the test treatment relative to placebo
#           proc         =c("restricted","unrestricted")
#                        "restricted": the variance sigma^2 is estimated
#                            by restricted ML to the null-hypothesis
#                        "unrestricted": the variance sigma^2 is estimated
#                            by unrestricted ML
#-----------------------------------------------------------------------------
#   OUTPUT: proc         Variance estimating procedure
#           teststatistic   Value of the teststatistic T
#           p.value         To the teststatistic corresponding p-value
#
#=============================================================================
RET.Pois.MinusId <- function(x.t,x.r,x.p,
                             n.t,n.r,n.p,
                             delta,
                             proc=c("restricted","unrestricted"))
{
  # MLE's ---------------------------------------------------------------------
  la.ml.t <- sum(x.t)/n.t
  la.ml.r <- sum(x.r)/n.r
  la.ml.p <- sum(x.p)/n.p


  # Estimation of the variance sigma^2 ----------------------------------------
  if (proc=="restricted"){
    # Computation of the restricted MLE's -----------------------------------

      # log likelihood in one group -----------------------------------------
      l.one <- function(lambda,k,n){
          k*log(lambda) -n*lambda
      }

    # minus restricted log likelihood ---------------------------------------
      ln.res <- function(par,x.t,x.r,x.p,n.t,n.r,n.p,d){
          res <- -l.one(d*par[1]+(1-d)*par[2],x.t,n.t) -
              l.one(par[1],x.r,n.r) - l.one(par[2],x.p,n.p)
          return(res)
```

```
                }
60
          res <- optim(c(la.ml.r,la.ml.p),ln.res,x.t=x.t,x.r=x.r,x.p=x.p,n.t=n.t,
62                      n.r=n.r,n.p=n.p,d=delta,method = "L-BFGS-B",lower=c(0,0))

64      la.rml.r <- res$par[1]
        la.rml.p <- res$par[2]
66      la.rml.t <- delta*la.rml.r + (1-delta)*la.rml.p

68      sigma <- sqrt(la.rml.t/n.t + delta^2*la.rml.r/n.r +
                      (1-delta)^2*la.rml.p/n.p)
70    }
      if (proc=="unrestricted"){
72      sigma <- sqrt(la.ml.t/n.t + delta^2*la.ml.r/n.r+
                      (1-delta)^2*la.ml.p/n.p)
74    }
    # Teststatistic ——————————————————————————————————————
76    teststatistic <- (-la.ml.t + delta*la.ml.r + (1-delta)*la.ml.p)/sigma
    #——————————————————————————————————————————————————————
78    result <- cbind(proc,round(teststatistic,4),round(1-pnorm(teststatistic),4))
      colnames(result) <- c("proc","teststatistic","p.value")
80    result <- as.data.frame(result)
      return(result)
82 }


84 #——————————————————————————————————————————————————————————
    #   EXAMPLE:
86 #   Performing the RET for an example in the treatment of epilepsy:
    #
88 #      Treatment       Total no. of seizures       No. of patients
    #    —————————————————————————————————————————————————
90 #        T(est)                  288                       18
    #      R(eference)              295                       18
92 #       P(lacebo)                338                       18
    #
94 #    (see Section 2.2, Table 2.2)
    #——————————————————————————————————————————————————————————
96 RET.Pois.MinusId(x.t=288,x.r=295,x.p=338,n.t=18,n.r=18,n.p=18,delta=0.5,
                                              proc="restricted")
98 RET.Pois.MinusId(x.t=288,x.r=295,x.p=338,n.t=18,n.r=18,n.p=18,delta=0.5,
                                              proc="unrestricted")

100
    #——————————————————————————————————————————————————————————
102 # OUTPUT:
    #
104 # > RET.Pois.MinusId(x.t=288,x.r=295,x.p=338,n.t=18,n.r=18,n.p=18,delta=0.5,
    # +                                            proc="restricted")
106 #         proc teststatistic p.value
    # 1 restricted          1.3281   0.0921
108 # > RET.Pois.MinusId(x.t=288,x.r=295,x.p=338,n.t=18,n.r=18,n.p=18,delta=0.5,
    # +                                            proc="unrestricted")
110 #         proc teststatistic p.value
    # 1 unrestricted         1.3491   0.0886
112 #——————————————————————————————————————————————————————————
    ############################################################################
114
    #——————————————————————————————————————————————————————————
116 #   FUNCTION: RET.Pois.MinusId.OptAlloc
    # ————————————————————————————————————————————————————
118 #   INPUT:    la.t          Success probability in the test group
```

```
      #            la.r        Success probability in the reference group
120   #            la.p        Success probability in the placebo group
      #            delta       Non−inferiority margin:
122   #                        Fraction of active controll effect which should be
      #                        retained by the test treatment relative to placebo
124   # ──────────────────────────────────────────────────────
      #   OUTPUT:              Optimal sample allocation:
126   #            w.t         Relative proportion of the test group        |
      #            w.r         Relative proportion of the reference group   | rounded on
128   #            w.p         Relative proportion of the placebo group     | 3 digits
      #
130   #════════════════════════════════════════════════════════
      RET.Pois.MinusId.OptAlloc <- function(la.t,la.r,la.p,delta){
132
        w.t <- sqrt(la.t)
134     w.r <- delta*sqrt(la.r)
        w.p <- abs(1−delta)*sqrt(la.p)
136
        result <- round(cbind(w.t,w.r,w.p)/(w.t+w.r+w.p),3)
138     colnames(result) <- c("w.t","w.r","w.p")
        result <- as.data.frame(result)
140
        return(result)
142   }


144   #════════════════════════════════════════════════════════
      #   EXAMPLE
146   #
      #   Optimal allocation for the RET under the alternative
148   #   la.t=la.r=10,  la.p=20 and delta=70%
      #════════════════════════════════════════════════════════
150   RET.Pois.MinusId.OptAlloc(la.t=10,la.r=10,la.p=20,delta=0.7)


152   #──────────────────────────────────────────────────────
      # OUTPUT:
154   # > RET.Pois.MinusId.OptAlloc(la.t=10,la.r=10,la.p=20,delta=0.7)
      #     w.t   w.r w.p
156   # 1 0.471  0.33 0.2
      #──────────────────────────────────────────────────────
158   ############################################################################


160   #════════════════════════════════════════════════════════
      #   FUNCTION: RET.Pois.MinusId.Samplesize
162   # ──────────────────────────────────────────────────────
      #   INPUT:    la.t        Lambda in the test group
164   #            la.r        Lambda in the reference group
      #            la.p        Lambda in the placebo group
166   #            w.t         Relative sample allocation to the test group (n.t/n)
      #            w.r         Relative sample allocation to the reference group (n.r/n)
168   #            delta       Non−inferiority margin:
      #                        Fraction of active controll effect which should be
170   #                        retained by the test treatment relative to placebo
      #            level       Significance level
172   #            pow         Aspired power (one minus the type II−error)
      #            proc        =c("restricted","unrestricted")
174   #                        "restricted": the variance sigma^2 will be estimated
      #                              by restricted ML to the null−hypothesis
176   #                        "unrestricted": the variance sigma^2 will be estimated
      #                              by unrestricted ML
178   # ──────────────────────────────────────────────────────
```

```
   #  OUTPUT:   proc      Variance estimating procedure, which will be used
180 #
   #                      Required sample sizes:
182 #           n          Total required sample size (n.t+n.r+n.p)
   #           n.t        Required samples for the test group (rounded)
184 #           n.r        Required samples for the refernce group (rounded)
   #           n.p        Required samples for the placebo group (rounded)
186 #
   #==============================================================================
188 RET.Pois.MinusId.Samplesize <- function(la.t,la.r,la.p,
                                          w.t,w.r,
190                                       delta,level,pow,
                                          proc=c("restricted","unrestricted"))
192 {
     if (proc=="restricted"){
194    #--- Kullback-Leibler divergence ----------------------------------
       KL <- function(la.0, la)
196      {
         la - la.0 + la.0 * log(la.0/la)
198      }
     #--- weighted Kullback-Leibler divergence -------------------------------
200    weighted.KL <- function(par,la.t.0,la.r.0,la.p.0,w.t,w.r,d)
       {
202      la.r <- par[1]
         la.p <- par[2]
204
         w.KL <- w.t*KL(la.t.0,d*la.r+(1-d)*la.p) +
206            w.r*KL(la.r.0,la.r) + (1-w.r-w.t)*KL(la.p.0,la.p)
         return(w.KL)
208      }
     #---------------------------------------------------------------------------
210    # KL-minimizer --------------------------------------------
       par <- optim(c(la.r,la.p), weighted.KL,
212               la.t.0=la.t,la.r.0=la.r,la.p.0=la.p,
                  w.t=w.t,w.r=w.r,d=delta)$par
214    la.rml.r <- par[1]
       la.rml.p <- par[2]
216    la.rml.t <- delta*la.rml.r + (1-delta)*la.rml.p
218    sigma.sq.rml <-  la.rml.t/w.t + (delta^2)*la.rml.r/w.r +
                        (1-delta)^2 * la.rml.p/(1-w.r-w.t)
220
       sigma.sq     <-  la.t/w.t + (delta^2)*la.r/w.r +
222                     (1-delta)^2 * la.p/(1-w.r-w.t)
224    eta <- -la.t+delta*la.r+(1-delta)*la.p
226    z.alpha <- qnorm(level)
       z.beta  <- qnorm(1-pow)
228
       n <- (z.alpha*sqrt(sigma.sq.rml)+z.beta*sqrt(sigma.sq))^2/eta^2
230  }
     if (proc=="unrestricted"){
232
       sigma.sq     <-  la.t/w.t + (delta^2)*la.r/w.r +
234                     (1-delta)^2 * la.p/(1-w.r-w.t)
236    eta <- -la.t+delta*la.r+(1-delta)*la.p
238    z.alpha <- qnorm(level)
```

```
        z.beta   <- qnorm(1-pow)
240
        n <- (z.alpha*sqrt(sigma.sq)+z.beta*sqrt(sigma.sq))^2/eta^2
242    }
      # rounding the sample sizes in each group ————————————————
244    n.t <- round(n*w.t)
      n.r <- round(n*w.r)
246    n.p <- round(n*(1-w.t-w.r))

248    result <- cbind(proc,n.t+n.r+n.p,n.t,n.r,n.p)

250    colnames(result) <- c("proc","n","n.t","n.r","n.p")
      result <- as.data.frame(result)
252
      return(result)
254 }


256 #══════════════════════════════════════════════════════════
    #   EXAMPLE
258 #
    # Determine the minimal required sample size for the alternative
260 # la.t=la.r=10, la.p=20 and delta=70% under the optimal sample
    # allocation (w.t,w.r,w.p)=(0.471,0.330,0.200) computed with
262 # 'RET.OptAlloc.Pois.MinusId()' when the variance sigma^2 will
    # be estimated with restriction to null hypothesis.
264 #══════════════════════════════════════════════════════════
    RET.Pois.MinusId.Samplesize(la.t=10,la.r=10,la.p=20,w.t=0.471,w.r=0.33,
266                                        delta=0.7,level=0.05,pow=0.8,
                                         proc="restricted")
268
    #————————————————————————————————————————————————————
270 # OUTPUT:
    # > RET.Pois.MinusId.Samplesize(la.t=10,la.r=10,la.p=20,w.t=0.471,w.r=0.33,
272 # +                                      delta=0.7,level=0.05,pow=0.8,
    # +                                      proc="restricted")
274 #          proc   n n.t n.r n.p
    # 1 restricted 31   15   10    6
276 #————————————————————————————————————————————————————
    ##############################################################################
```

A.3   Normally distributed endpoints

```
###############################################################################
2  #————————————————————————————————————————————————————————————————————#
   #                                                                      #
4  # NORMALLY DISTRIBUTED ENDPOINTS                                       #
   #                                                                      #
6  # For a detailed discussion see Pigeot et al. (2003).                  #
   #                                                                      #
8  ###############################################################################

10 #=====================================================================
   #   FUNCTION: RET.Norm.Id
12 # —————————————————————————————————————————————————————————————————
   #   INPUT:  x.t          Vector with observations from the test group
14 #           x.r          Vector with observations from reference group
   #           x.p          Vector with observations from placebo group
16 #           delta        Non−inferiority margin:
   #                        Fraction of active controll effect which should be
18 #                        retained by the test treatment relative to placebo
   # —————————————————————————————————————————————————————————————————
20 #   OUTPUT: teststatistic    Value of the teststatistic T
   #           p.value          To the teststatistic corresponding p−value
22 #
   #=====================================================================
24 RET.Norm.Id <− function(x.t,x.r,x.p,delta)
   {
26 # MLE's ———————————————————————————————————————————————————————————
   mu.ml.t <− mean(x.t)
28 mu.ml.r <− mean(x.r)
   mu.ml.p <− mean(x.p)

30
   n.t <− length(x.t)
32 n.r <− length(x.r)
   n.p <− length(x.p)
34 n <− n.t+n.r+n.p

36 sigma <− sqrt( (1/n.t + delta^2/n.r + (1−delta)^2/n.p)*
                  ((n.t−1)*sd(x.t)^2+(n.r−1)*sd(x.r)^2+(n.p−1)*sd(x.p)^2)/(n−3))
38
   # Teststatistic —————————————————————————————————————————————————
40 teststatistic <− (mu.ml.t − delta*mu.ml.r−(1−delta)*mu.ml.p)/sigma

42 result <− cbind(round(teststatistic,4),round(1−pnorm(teststatistic),4))
   colnames(result) <− c("teststatistic","p.value")
44 result <− as.data.frame(result)
   return(result)
46 }


48 #=====================================================================
   #   EXAMPLE
50 #=====================================================================
   # ——— Generating some data ———————————————————————————————————————
52 set.seed(666)
   x.t <− round(rnorm(100,10,1),2)   # 100 rounded samples from N(10,1)
54 x.r <− round(rnorm(80,10,1),2)    #  80 rounded samples from N(10,1)
   x.p <− round(rnorm(70,9,1),2)     #  70 rounded samples from N(9,1)
56
   # Analyzing data ———————————————————————————————————————————————————
58 RET.Norm.Id(x.t,x.r,x.p,delta=0.7)
```

```
60  #—————————————————————————————————————————————
    #  OUTPUT:
62  #  > RET.Norm.Id(x.t,x.r,x.p,delta=0.7)
    #     teststatistic  p.value
64  #  1          1.924    0.0272
    #—————————————————————————————————————————————
66  ############################################################################

68  #————————————————————————————————————————————————
    #   FUNCTION: RET.Norm.Id.Samplesize
70  # ———————————————————————————————————————————————
    #   INPUT:    mu.t       Expectation of the test group
72  #             mu.r       Expectation of the reference group
    #             mu.p       Expectation of the placebo group
74  #             p.sigma     Standard deviation of the groups k=t,r,p, which
    #                        is presumed to be homogeneous
76  #             w.t        Relative sample allocation to the test group (n.t/n)
    #             w.r        Relative sample allocation to the reference group (n.r/n)
78  #             delta      Non−inferiority margin:
    #                        Fraction of active controll effect which should be
80  #                        retained by the test treatment relative to placebo
    #             level      Significance level
82  #             pow        Aspired power (one minus the type II−error)
    # ———————————————————————————————————————————————
84  #   OUTPUT:
    #                        Required sample sizes:
86  #             n          Total required sample size (n.t+n.r+n.p)
    #             n.t        Required samples for the test group (rounded)
88  #             n.r        Required samples for the refernce group (rounded)
    #             n.p        Required samples for the placebo group (rounded)
90  #
    ############################################################################
92  RET.Norm.Id.Samplesize <− function(mu.t,mu.r,mu.p,sigma,
                                        w.t,w.r,delta,level,pow)
94  {
        sigma.sq <−  (1/w.t + delta^2/w.r + (1−delta)^2/(1−w.r−w.t))*sigma^2
96
        eta <− mu.t−delta*mu.r−(1−delta)*mu.p
98
        z.alpha <− qnorm(level)
100     z.beta  <− qnorm(1−pow)
102     n <− (z.alpha*sqrt(sigma.sq)+z.beta*sqrt(sigma.sq))^2/eta^2
104     # rounding the sample sizes in each group ——————————————————
        n.t <− round(n*w.t)
106     n.r <− round(n*w.r)
        n.p <− round(n*(1−w.t−w.r))
108
        result <− cbind(n.t+n.r+n.p,n.t,n.r,n.p)
110
        colnames(result) <− c("n","n.t","n.r","n.p")
112     result <− as.data.frame(result)
114     return(result)
    }
116
    #—————————————————————————————————————————————
118 #   EXAMPLE
```

```
      #
120   # Determine the minimal required sample size for the alternative
      # mu.t=mu.r=10, mu.p=9 and delta=80% under the optimal sample
122   # allocation (w.t,w.r,w.p)=(0.5,0.4,0.1) and a variance of 1.
      #========================================================================
124   RET.Norm.Id.Samplesize(mu.t=10,mu.r=10,mu.p=9,sigma=1,w.t=0.5,w.r=0.4,
                                        delta=0.8,level=0.05,pow=0.8)
126
      #------------------------------------------------------------------------
128   # OUTPUT:
      # > RET.Norm.Id.Samplesize(mu.t=10,mu.r=10,mu.p=9,sigma=1,w.t=0.5,w.r=0.4,
130   # +                                   delta=0.8,level=0.05,pow=0.8)
      #      n n.t n.r n.p
132   # 1 618 309 247  62
      #------------------------------------------------------------------------
134   ########################################################################
```

A.4   Censored, exponentially distributed endpoints

```
1  #####################################################################################
   #——————————————————————————————————————————————————————————————————————————————#
3  #                                                                                #
   # CENSORED, EXPONENTIALLY DISTRIBUTED ENDPOINTS                                  #
5  #                                                                                #
   # For a detailed discussion see Mielke et al. (2008).                           #
7  #                                                                                #
   #####################################################################################
9
   #————————————————————————————————————————————————————————————————————————————————
11 #   FUNCTION: RET.Exp.MinusLog
   # ——————————————————————————————————————————————————————————————————————————————
13 #   INPUT: x.t           Observations of the test group
   #          x.r           Observations of the reference group
15 #          x.p           Observations of the placebo group
   #                 ———————————————————————————————————————————
17 #          where x.k is a matrix with one single observation
   #          in each row and two columns, the observed value in
19 #          the first column and the censoring indicator in
   #          the second column
21 #                 ———————————————————————————————————————————
   #          delta         Non—inferiority margin:
23 #                         Fraction of active controll effect which should be
   #                         retained by the test treatment relative to placebo
25 # ——————————————————————————————————————————————————————————————————————————————
   #   OUTPUT: teststatistic   Value of the teststatistic T
27 #           p.value         To the teststatistic corresponding p—value
   #
29 #————————————————————————————————————————————————————————————————————————————————
   RET.Exp.MinusLog <- function(x.t,x.r,x.p,delta)
31 {
   # number of uncensored observations ————————————————————————————————————————————
33  cen_t <- sum(x.t[,2])
    cen_r <- sum(x.r[,2])
35  cen_p <- sum(x.p[,2])

37  # MLE's ————————————————————————————————————————————————————————————————————————
    la.ml.t <- sum(x.t[,1])/cen_t
39  la.ml.r <- sum(x.r[,1])/cen_r
    la.ml.p <- sum(x.p[,1])/cen_p
41
    #——————————————————————————————————————————————————————————————————————————————
43  sigma <- sqrt(1/cen_t + delta^2/cen_r + (delta-1)^2/cen_p)

45  # Teststatistic ————————————————————————————————————————————————————————————————
    teststatistic <- (-log(la.ml.t) + delta*log(la.ml.r) +
47                    (1-delta)*log(la.ml.p))/sigma

49  #——————————————————————————————————————————————————————————————————————————————
    result <- cbind(round(teststatistic,4),round(1-pnorm(teststatistic),4))
51  colnames(result) <- c("teststatistic","p.value")
    result <- as.data.frame(result)
53  return(result)
   }
55
   #————————————————————————————————————————————————————————————————————————————————
57 #   EXAMPLE
   #————————————————————————————————————————————————————————————————————————————————
```

```
59 # ——— Generating some data ——————————————————————————————————
   # Number of samples
61 n.t  <- 200
   n.r  <- 150
63 n.p  <- 100

65 set.seed(666)
   t.t <- rexp(n.t,1/10)
67 t.r <- rexp(n.r,1/10)
   t.p <- rexp(n.p,1/15)
69 u.t <- rexp(n.t,1/10)      #
   u.r <- rexp(n.r,1/10)      # censoring variables
71 u.p <- rexp(n.p,1/15)      #

73 x.t <- cbind(apply(cbind(t.t,u.t),1,min),as.numeric(t.t<=u.t))
   x.r <- cbind(apply(cbind(t.r,u.r),1,min),as.numeric(t.r<=u.r))
75 x.p <- cbind(apply(cbind(t.p,u.p),1,min),as.numeric(t.p<=u.p))

77 # Analyzing data ——————————————————————————————————————
   RET.Exp.MinusLog(x.t,x.r,x.p,delta=0.7)

79
   #——————————————————————————————————————————————————————
81 # OUTPUT:
   # > RET.Exp.MinusLog(x.t,x.r,x.p,delta=0.7)
83 #     teststatistic p.value
   # 1         1.7341   0.0415
85 #——————————————————————————————————————————————————————
   ##############################################################################

87
   #——————————————————————————————————————————————————————
89 #   FUNCTION: RET.Exp.MinusLog.Samplesize
   # ——————————————————————————————————————————————————————
91 #   INPUT:    par.t    Parameter of the test group
   #            par.r    Parameter of the reference group
93 #            par.p    Parameter of the placebo group
   #            ———————————————————————————————————————————
95 #            where each parameter par.k consists of the pair
   #            "(la,prob)", where "la" represents the parameter
97 #            of the exponential distribution (with expectation la)
   #            and "prob" represents the probability for an
99 #            uncensored observation.
   #            ———————————————————————————————————————————
101 #           w.t      Relative sample allocation to the test group (n.t/n)
    #           w.r      Relative sample allocation to the reference group (n.r/n)
103 #           delta    Non-inferiority margin:
    #                    Fraction of active controll effect which should be
105 #                    retained by the test treatment relative to placebo
    #           level    Significance level
107 #           pow      Aspired Power (one minus the type II-error)
    #
109 # ——————————————————————————————————————————————————————
    #   OUTPUT:
111 #                    Required sample sizes:
    #           n        Total required sample size (n.t+n.r+n.p)
113 #           n.t      Required samples for the test group (rounded)
    #           n.r      Required samples for the refernce group (rounded)
115 #           n.p      Required samples for the placebo group (rounded)
    #
117 ##############################################################################
    RET.Exp.MinusLog.Samplesize <- function(par.t,par.r,par.p,
```

```r
119                                               w.t,w.r,delta,level,pow)
    {
121    la.t      <-   par.t[1]
       la.r      <-   par.r[1]
123    la.p      <-   par.p[1]
       prob.t    <-   par.t[2]
125    prob.r    <-   par.r[2]
       prob.p    <-   par.p[2]
127
       sigma.sq <-    1/(w.t*prob.t) + delta^2/(w.r*prob.r) +
129                   (1-delta)^2/((1-w.r-w.t)*prob.p)

131    eta <- -log(la.t)+delta*log(la.r)+(1-delta)*log(la.p)

133    z.alpha <- qnorm(level)
       z.beta  <- qnorm(1-pow)
135
       n <- (z.alpha*sqrt(sigma.sq)+z.beta*sqrt(sigma.sq))^2/eta^2
137
       # rounding the sample sizes in each group ————————————————
139    n.t <- round(n*w.t)
       n.r <- round(n*w.r)
141    n.p <- round(n*(1-w.t-w.r))

143    result <- cbind(n.t+n.r+n.p,n.t,n.r,n.p)

145    colnames(result) <- c("n","n.t","n.r","n.p")
       result <- as.data.frame(result)
147
       return(result)
149 }


151 #═══════════════════════════════════════════════════════════
    #   EXAMPLE
153 #═══════════════════════════════════════════════════════════
    par.t <- c(10,0.8)
155 par.r <- c(10,0.8)
    par.p <- c(20,0.8)
157
    RET.Exp.MinusLog.Samplesize(par.t,par.r,par.p,w.t=0.5,w.r=0.4,
159                               delta=0.8,level=0.05,pow=0.8)

161 #——————————————————————————————————————————————————————————————
    # OUTPUT:
163 # > RET.Exp.MinusLog.Samplesize(par.t,par.r,par.p,w.t=0.5,w.r=0.4,
    # +                             delta=0.8,level=0.05,pow=0.8)
165 #       n n.t n.r n.p
    # 1  1608 804 643 161
167 #——————————————————————————————————————————————————————————————
    ############################################################################
```

Curriculum Vitae

Matthias Mielke

born 24. March 1981 in Uslar
Single, German

1987 – 2000	Schooling *Abitur* at the *Theodor-Heuss-Gymnasium Göttingen*
2000 – 2001	Civil service
2001 – 2007	Diploma studies in mathematics and business studies (minor) *Faculty of Mathematics, University of Göttingen* diploma thesis: *Die asymptotische Verteilung des Likelihood-Quotienten-Tests für allgemeine Hypothesenräume* supervised by *Prof. Dr. Axel Munk*
2007 – 2010	Ph.D. studies in mathematics *Institute for Mathematical Stochastics, University of Göttingen* supervised by *Prof. Dr. Axel Munk*
2007 – 2010	Member of the Ph.D. Programm "Applied Statistics and Empirical Methods" *Centre for Statistics, University of Göttingen*
02 – 09/2007	"Georg-Lichtenberg"-scholar
since 10/2007	Assistant at the Institute for Mathematical Stochastics, University of Göttingen