Final Summary Report

Grant agreement number: Health-F5-2013-603160

Project acronym: ASTERIX

Project title: **A**dvances in **S**mall **T**rials d**E**sign for **R**egulatory **I**nnovation and e**X**cellence.

Start date project: October 1st, 2013

Period covered: October 1st, 2013 to September 30th, 2017

Name of the scientific representative of the project's coordinator, title and organisation:

Prof. dr. Kit C.B. Roes

University Medical Center Utrecht (PIC 999915189)

Heidelberglaan 100

3508 GA Utrecht, Netherlands

Tel: +31 88 7568295

e-mail: k.c.b.roes@umcutrecht.nl

website: www.asterix-fp7.eu

Table of Contents

1.1 FINAL PUBLISHABLE SUMMARY REPORT 3

1.1.1 EXECUTIVE SUMMARY 3
1.1.2 SUMMARY DESCRIPTION OF PROJECT CONTEXT AND OBJECTIVES 4
1.1.3 DESCRIPTION OF THE MAIN S&T RESULTS/FOREGROUND 7
1.1.3.1 Management and collaboration 7
1.1.3.2 Single trial methodology 9
1.1.3.3 Series of trials methodology 14
1.1.3.4 Patient perspective 20
1.1.3.5 Level of evidence 24
1.1.3.6 Framework, validation and regulatory decision making 27
1.1.4 POTENTIAL IMPACT, MAIN DISSEMINATION ACTIVITIES AND EXPLOITATION OF RESULTS 33
1.1.4.1 Potential Impact 33
1.1.4.2 Main dissemination activities 38
1.1.4.3 Exploitation of results 42
REFERENCES 43

1.1 Final Publishable Summary Report

1.1.1 Executive summary

An estimated 30 million patients in Europe suffer from one of the more than 7000 known rare diseases. For many of these potentially very debilitating diseases a treatment still has to be found. Having 1000 new treatments available in the next decade has been announced in Nature as one of the goals of the International Rare Diseases Research Consortium. This ambitious goal requires the development and evaluation of innovative drugs in small populations. To advance drug development and decision making in rare diseases, the efficiency, reliability and value of clinical trials used for evaluation of a new drug need to be improved. The ASTERIX project aimed to do this by optimizing methodology in rare diseases in a multi-stakeholder team, focusing on (1) statistical methodology for design and analysis, (2) incorporation of the patient perspective in design and outcomes and (3) uptake in practice and regulatory guidance.

ASTERIX progressed with new **statistical methodology**, and deepened insight in how methodology impacts assessment of evidence. The achievements include the following.

- Guidance on stratification and minimization in clinical trials for rare diseases.
- New methods to make optimal use of multiple endpoints.
- New and improved adaptive designs, tailored to settings for rare diseases.
- New methods to incorporate information from previous trials in the design and analysis of trials for rare diseases.
- Thorough understanding and recommendations for meta-analyses in case of a small number of small trials.
- Thorough understanding and development of evidentiary standards for individual trials and drug development strategies, including the importance of randomisation and alternative strategies for exceptional circumstances.

Importantly, ASTERIX successfully included a *Patient Think Tank* to guide the entire project from **a patient perspective** – and hence we "practiced what we preached". In addition, the specific patient focused research results included:

- Guidance and checklist to optimise use of patient registries to inform design.
- Practical model (POWER) to include patient's preferences in the weighing of outcomes.
- Sound statistical and conceptual basis for Goal Attainment Scaling to serve as outcome instrument that reflects the inherent heterogeneity in rare disease trials.
- Ethical aspects specific to rare disease clinical research were explored, resulting in an "ethical benchmark".

Finally, to guide actual **uptake in practice and support regulatory guidance,** the novel methods were evaluated against orphan drug development plans (based on European Public Assessment Reports), using a comprehensive medical condition clustering framework.
Advices and novel methods developed within ASTERIX have already been applied in clinical trials and drug development plans. To reach full **impact**, the scale of application of the developed innovations needs to be broadened. To achieve this, some of the innovative methods require both further development as well as a shift in mind-set. Long term sustainment of the innovations was ensured by making PhD projects the core of the program: the "next generation" of clinical trial methodology experts was educated, ready to support trials in rare diseases.

For more information: www.asterix-fp7.eu

1.1.2 Summary description of project context and objectives

Unmet need for novel drugs to treat rare diseases

The European legislation on orphan medicinal products [Regulation (EC) No 141/2000] emphasises that patients suffering from rare conditions should be entitled to the same quality of treatment as other patients. The combined clinical and socio-economic impact of rare diseases is huge. An estimated 30 million European patients suffer from a rare disease, and more than 7000 rare diseases are known, many of them chronic, potentially very disabling and typically affecting children. To stimulate research and development of treatments for rare diseases the European Legislation provides incentives for manufacturers who develop drugs for rare life-threatening or chronically debilitating diseases. Since 2000, more than 850 orphan drug designations have been granted by the European Commission. Between 2000 and 2013, 157 received an opinion from the Committee for Medicinal Products for Human use (CHMP), which was positive for 104. These obtained market authorisations (Hofer et al., 2017). By the end of 2016, 128 market authorisations were granted for orphan designated drugs. This could be seen as success for the stimulating legislation. However, as follows from the research in Hofer et al., essential issues remain regarding clinical trial design, choice of endpoint and robustness of the methodology that still make failure at registration for drugs for rare medical conditions more likely than for common conditions (from Hofer et al.: 34% versus 21% negative opinions, respectively).

Most importantly, this slows access to new treatments for patients suffering from a rare disease. The International Rare Diseases Research Consortium (IRDiRC) previously set a goal to contribute to the development of 200 new medicinal treatments for rare diseases by 2020. The present achievement (December 2016, source: IRDiRC website) is that 222 new products for rare diseases obtained market authorisation across the US and Europe. Although this exceeds the objective, it is still a very modest amount against the over 7000 known rare diseases. Therefore, IRDiRC is considering to propose new goals, including having 1000 new treatments for rare diseases available through the next decade (Austin and Dawkins, 2017). This is an ambitious goal, that firstly requires innovative drugs. To achieve such ambitious goals, many drugs will need evaluation through clinical trials, performed in small populations. Arguably, efficient and effective clinical research is subsequently the most important development phase to improve, given its associated costs and risks. It is, however, intrinsically difficult to demonstrate efficacy and positive benefit-risk with the current (evidence and regulatory) standards and methodology in clinical trials for rare diseases, for which the prevalence prohibits recruiting the numbers of patients needed.

Efficiency and reliability of clinical trials are indeed strongly (but not solely) determined by the statistical design and analyses applied. These determine, together with the quality of endpoint assessments, the power and sample size of trials and the precision with which treatment effects can be determined. In addition, relevance of trials and outcomes to patients is a strong determinant of the ultimate value of trials for benefit-risk assessment. Finally, improved methodology can only have impact if it has demonstrated added value in practice and supports regulatory decision making.

Thus, ASTERIX decided to focus on progress in clinical research for new treatments for *rare diseases*. The vision of ASTERIX is that such progress can be best made, by advancing *in coherence*: (1) statistical methodology for design and analysis, (2) incorporation of the patient perspective in design and outcomes and (3) uptake in practice and regulatory guidance.

Aligned with this vision, ASTERIX included a diversity of researchers and stakeholders to be able to tackle these challenges, spurring interactions between researchers from different fields, patient representatives, industry and regulators. Unique for

a methodological research project was inclusion of an active Patient Think Tank in the project and at several plenary scientific meetings.

Statistical methodology for design and analysis

There is a vast body of statistical methodology for design and analysis of clinical trials. A framework for small populations was suggested (Gupta et al., 2011) that systemically advises a design depending on the duration of clinical effect, required follow-up versus accrual time, anticipated sample sizes, difficulties of retention of patients and prior level of confidence in effect of treatment. These include some designs specific for small populations (n-of-1, response adaptive allocation design). A number of these trial designs are only applicable under restrictive assumptions, that will not hold for many rare diseases. Many recent improvements in clinical trial design, such as sequential and adaptive designs, are in principle also applicable to small populations, and may require smaller numbers of patients than conventional designs to obtain conclusive evidence. Characteristics of these design are typically only well-understood when applied to large samples. No accepted solutions exist for these designs in case of truly small samples, multiple endpoints or individualised patient outcomes that typically become relevant if patient perspectives are included. Meta-analysis of (a sequence) of trials is an available method; previous reluctance against re-use of clinical trials data in a meta-analysis has been addressed and the role of meta-analysis in general drug development has been carefully explored, but also here typically in the large sample situation. Issues such as heterogeneity between trials and incorporating studies with different designs have not been addressed when the number of trials and the information per trial is limited. These were identified as methodological gaps in translating more efficient trial designs to tailored solutions for clinical research in rare diseases.

Incorporation of patient perspectives in clinical trial designs

Many initiatives have increased awareness to include the patient perspectives in the design of clinical research. No systematic approach exists to date to incorporate this information in actual “hard-core” design and analysis. This is particularly relevant in rare diseases, since due to the small populations there is limited opportunity for replication and extension of research to separately address these perspectives. Key areas of improvement that will typically fit rare diseases are: (1) use of information and data from patient registries; (2) explicit elicitation of patient preferences for treatment outcomes as well as other design features and (3) development of patient centred (individualised) outcomes. The latter can have impact both through increased relevance as well as increased patient participation in trials. These were the three patient focused topics for ASTERIX. In addition it was realised that the rare disease setting might incur specific ethical issues for conducting clinical research, which was to be explored through appropriate qualitative research methods. The lack of patient involvement in design may lead to trials that do not optimally motivate patients to participate, while the recruitment potential is already limited. ASTERIX aimed to develop and test a practical approach to facilitate involvement in the design of rare disease trials.

Uptake in practice and regulatory guidance

More than 10 years have lapsed since the issue of Regulation(EC) No 141/2000, and substantial experience has been gained on the assessment of orphan medicinal product marketing applications. Also, a number of statistical methodologies potentially suited to clinical trials for rare diseases have become available. Most novel methods have been well characterised in large populations, but not in small trials with very limited possibility for replication. Lack of evidence on the robustness of novel methods may hamper further implementation. The EU legislation determines that market access to new drugs requires the same level of evidence regardless of whether they are intended for (very) rare or highly prevalent diseases (Regulation (EC) No 141/2000). While patients’ safety and best interests lead these provisions, they pose difficulties to developers that may discourage the research of new treatments for rare diseases and may delay the access to new or improved therapies

for rare disease populations (Roberts et al. 2011; Kesselheim et al., 2012; Jonsson and Bergh, 2012). The present standards of evidence, including required significance levels to demonstrate efficacy and acceptability of different designs (such as single arm trials, use of historical controls or non-inferiority), do not distinguish between rare or regular diseases. Nevertheless, actual decision making does seem to differ between rare and frequent diseases, but systematic guidance is lacking. ASTERIX aimed to improve this situation by preparing for guidance that is more specific than the current small populations guideline, and takes characteristics of the medical condition to be treated into account.

Concept of ASTERIX

ASTERIX aimed to optimize methodology for clinical research in rare diseases by coherently addressing the fundamental challenges addressed above. To this end, a multidisciplinary consortium consisting of statisticians, advisors to regulatory bodies, physicians, patient representatives and industry was formed. Importantly, ASTERIX aimed to develop and build upon a clustering of medical conditions to provide more specific guidance for choosing the best suited methodologies (see Figure 1). This framework, once developed, was to be validated against real life cases. These included case studies from the partner institutes as well as a systematic evaluation of regulatory dossiers.

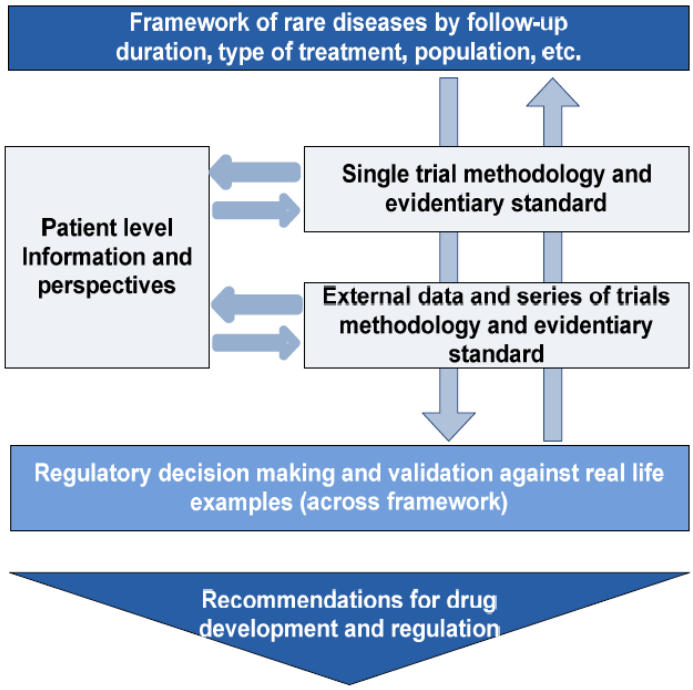


Figure 1. Concept of the ASTERIX project

Objectives of ASTERIX

The coherent concept of ASTERIX was translated into the following, more specific, objectives, stated as follows in the original project proposal.

Individual trials

- Provide guidance on stratification and minimization in clinical trials for rare diseases to maximize their reliability and efficiency.
- Improve the efficiency of clinical trials for rare diseases by methods to base licensing decisions on multiple endpoints and make the involved assumptions and risks transparent.

- Taylor guidance on adaptive designs to settings in drug development for rare diseases and develop adaptive clinical trial designs that make best use of information from multiple endpoints to improve the efficiency of clinical trials for rare diseases.
- Optimize evidentiary standards for individual trial designs.

Series of trials
- Extend sequential meta-analysis (SMA) methodology for prospective implementation in case of a small number of trials with small sample sizes and a single endpoint.
- Extend SMA methodology for application to multiple endpoints in rare diseases.
- Provide methods to incorporate historical information in the design and analysis of new trials for rare diseases.
- Develop adaptive methods to combine phase II and phase III trial designs in drug development for rare diseases.
- Optimize evidentiary standards for situations with evidence from multiple trials.

Patient perspective
- Optimize use of patient registries to inform design.
- Deliver methodology to include patient's preferences in the weighing of outcomes.
- Assessing the value of Goal Attainment Scaling to reflect the inherent heterogeneity in rare disease trials.
- Establish methods to facilitate patient involvement in trial design.

Framework and regulatory impact
- Produce a taxonomy of rare diseases to systematize applicability of study designs and statistical methods.
- Validate new statistical methods with real life clinical trial cases.
- Develop new regulatory models to decision making.
- Validate new statistical methods and regulatory models with real-life regulatory datasets.
- Test the acceptability of new methods and models by patients and regulators.
- Issue recommendations for new guidance.

1.1.3 Description of the main S&T results/foreground

1.1.3.1 Management and collaboration

From the outset, ASTERIX aimed to be a multidisciplinary and strongly collaborative project. Objectives were to be focused on new or improved methodology for clinical trials for new (drug) treatments for rare diseases. Statistical methodology is key in improving efficiency and reliability of clinical trial design, analysis and decision making. True success in the end depends on implementation and adoption, and this requires a multi-stakeholder perspective. Thus, researchers with different expertise participated (of course statisticians, but also epidemiologists, ethicists, clinical pharmacologists, health scientists) and stakeholders from diverse backgrounds – patients, regulatory, industry – were included in participative oversight. The **principles of managing** the project were as follows:

- Flat structure, no layers between overall management, researchers and all partners represented.
- Multidisciplinary management and advisory board oversight.
- Structure of work packages that fosters cross-partner collaboration.
- Shaped around PhD projects, combining research and education of the next generation of clinical trial scientists.

This resulted in the management structure depicted in Figure 2. The executive board interacted regularly through scheduled teleconferences and at all scientific meetings, very much informal to assure swift progress and decision making. To ensure progress, replacements would be present in absence of any partner representative.

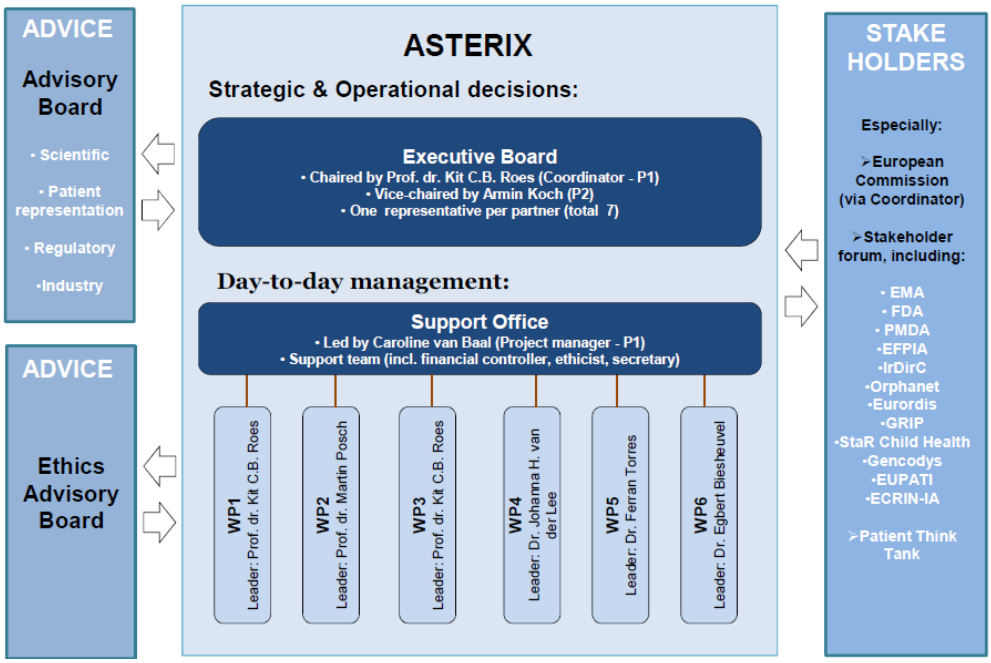
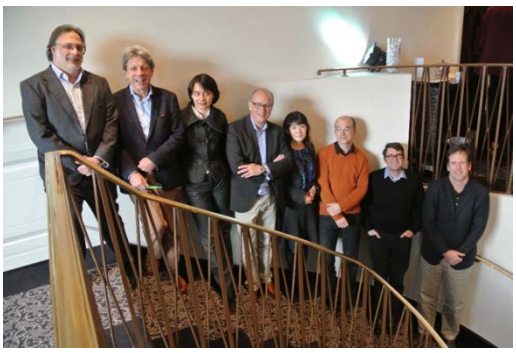


Figure 2. Project structure of ASTERIX.

Both a scientific as well as ethics advisory board was installed. The ethics advisory board was added on account of the patient interactions that were planned as part of the project and the research that was to be done on ethical aspects of new methodology. Both boards included excellent, international expertise – covering science, regulatory (including the chair of the Dutch Medicines Evaluation Board, Prof. H. Leufkens and a representative of the regulatory agency of Japan (PMDA), dr. Y. Ando) and the patient perspective. The advisory boards were invited to attend the ASTERIX consortium scientific meetings at least once per year, during which they also held their formal meetings. In the course of the project, they became highly **committed boards**, actively engaging in the discussions with the scientists.



Picture 1. The Executive Board and Advisory Boards.

ASTERIX arranged interactive sessions, e.g. poster sessions, during which board members interacted with individual scientists and their project to get a grip on the progress, challenges encountered and anticipated results. The boards appreciated this opportunity for

interaction, with more possibility for direct input then was possible with the usual way of presenting only overview documents and presentations. ASTERIX is most likely the first methodological project of this size, that aimed for structural involvement of patient representatives. This was illustrated at the meeting in Vienna in 2014 (jointly with the IDeAl and InSPiRE projects), where Peter Kapteijn from Inspire2Live presented the patient perspective for the small populations projects. In hindsight, we did not have a fully laid out approach on how to involve patients for a methodological project, with researchers that would not normally interact with patients professionally. Jointly with the **Patient Think Tank** we discovered and shaped a fruitful collaboration, resulting in a highly committed Patient Think Tank and (young) researchers that gained valuable insights into the contributions they can and should make to improve access for patients to new effective treatments. One of the highlights of the interaction included 3-minute pitches of the researchers to the Patient Think Tank, explaining their research and how this would benefit patients at the scientific meeting of ASTERIX in Barcelona in October 2015. The collaboration with the Patient Think Tank further resulted in patient representatives co-authoring the paper on patient involvement, and co-developing patient information leaflets to explain key (novel) clinical trial design and analysis features for patients that may enter clinical trials.

The management principles and focus on PhD projects with young scientists have resulted in a consortium that will continue to collaborate beyond the ASTERIX project. At the scientific meetings, the young researchers were able to build their network. In the course of the project exchanges of young researchers took place. Notably, one PhD student from the Academic Medical Center Amsterdam conducted part of her research at the Medical University in Vienna, one from Vienna spent time at the University Medical Center Utrecht and one from Utrecht at Hannover Medical School. Exchanges materialised in joint work and papers. At present, the next consortium meeting is being planned. The consortium is also deeply engaged in new initiatives that may help progress clinical research for rare diseases and children. In the spirit of the last four years, the collaboration has few borders: the three principal investigators of ASTERIX, IDeAl and InSPiRe are closely connected and foster the cross-project network that has emerged.

1.1.3.2 Single trial methodology

There is an abundance of methodology to improve the design and analysis of individual trials, often essentially aimed at increasing efficiency: extract more information from the same trial, increase the probability of success of an individual clinical trial and enabling the conduct of smaller trials. Yet, progress for clinical trials in truly small populations has proven difficult to achieve. ASTERIX targeted promising areas of improvement; some fundamental (such as randomisation and stratification), others tailoring advances to the challenges of small samples.

Stratification and minimization in clinical trials for rare diseases

Randomisation is the cornerstone of clinical trials, to avoid bias and provide robust results. In most clinical trials for drug development, randomisation is stratified – for logistical reasons related to medication supply, or to increase efficiency and (further) reduce bias in the presence of known prognostic factors. Our literature review on available methodology for randomisation and associated analyses showed that many papers contain recommendations that are based on tradition, but have neither been validated in trials with sample-size restrictions nor in trials under regular conditions. Among the stratified allocation methods discussed, stratified permuted block randomization and Pocock & Simon's marginal procedure are most commonly referred to and applied in clinical trials. The marginal procedures, in particular minimization, show promising (statistical) properties regarding balance and power, especially for studies with small sample sizes. There are also some drawbacks: little is known about the validity and correct use of analysis methods and the

behaviour of the marginal procedure in more extreme situations regarding allocation to strata (covariate patterns).

Based on an extensive simulation model, we examined under which circumstances stratified randomization procedures may lead to loss in power or type I error inflation. It is evident and common practice that the number of stratification factors needs to be limited by terms of balance and accidental bias. Limitations inherent to the analysis model, however, are commonly disregarded in design recommendations concerning allocation methodology. Small sample sizes impose stronger restrictions on analysis, especially in non-linear models such as logistic regression, than the balancing properties of common stratified allocation methods.

Recommendation

- The tolerable number of strata in a stratified randomization is limited both by the balancing properties of stratification and by the limitations of a stratified analysis. Therefore, like in frequent disease, only variables known to be prognostic should be incorporated in a stratified randomization.

Utilization of multiple endpoints

Typically, diseases and outcomes of treatments are multi-faceted. Hence, multiple endpoints are analysed in clinical trials to address the different aspects of a disease. Moreover, they can maximise the available information from a limited sample. A literature review on hypothesis tests for multiple endpoints with particular focus on small sample aspects was prepared comprising methods for the combination of endpoints as well as multiple testing procedures for individual endpoints (Ristl et al., 2017). Typically, control of the (family wise) type I error rate in trials with small sample sizes is achieved by replacement of purely asymptotic methods with tests based on t- or F-distributions and the use of resampling procedures.

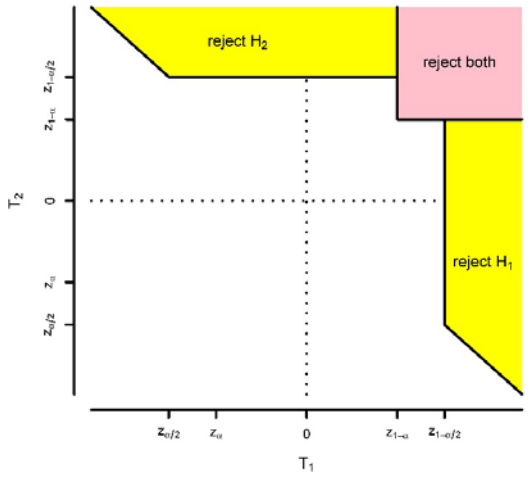


Figure 3. Test decisions with the diagonally trimmed Simes test, a fallback test for two co-primary endpoints. T_1 and T_2 are the test statistics for the two endpoints, respectively. $z_{1-\alpha}$ is the 1-α quantile of the common marginal null distribution of either test statistics. The rejection region for the classic co-primary endpoint test corresponds to the red region. With the fallback test, additional rejection rules corresponding to the yellow region are added. The type I error rate is controlled if T_1 and T_2 are multivariate normal or multivariate t-distributed (Ristl et al., 2016).

In complex diseases, efficacy needs to be demonstrated in two or more co-primary endpoints to show a clinical benefit. In the standard hypothesis testing approach, efficacy in all endpoints is concluded if all tests are significant at the overall significance level. However, in this setting no conclusion of efficacy can be drawn, *if some*, but not all endpoints show an effect. This is somewhat counterintuitive, and presents potential waste of clinical trial

information. We proposed a fallback test (Figure 3), that also allows for conclusions on individual endpoints even if the main study aim to show efficacy in all endpoints was not achieved. This test uniformly improves the standard testing procedures (Ristl et al., 2016).

When testing hypotheses for binary endpoints, such as responder yes/no or 30 day mortality, exact tests allow for strict control of the type I error rate in small clinical trials. However, these tests are conservative due to the discreteness of the test statistics. In a multivariate setting, conservatism of exact tests may be greatly reduced by studying the multivariate joint distribution of discrete test statistics and defining a suitable multivariate rejection region. In (Ristl et al., 2016) an optimization framework is proposed to find multivariate rejection regions with certain optimality properties, e.g. maximal exhaustion of the nominal type I error rate or maximal power under a pre-specified point alternative. The methodology was implemented in the open source R-package *multfisher* which is available on CRAN.

Furthermore, we developed a powerful hypothesis testing framework for clinical trials with multiple endpoints of different scale (e.g. metric, binary and count data) that are repeatedly observed within patients (Ristl et al., 2017). The method is based on multiple generalized estimating equation models, taking into account dependencies within the same subject as well as the correlation between endpoints. Simultaneous confidence intervals and hypothesis tests for linear contrasts of regression coefficients of the multiple marginal models are derived and special emphasis is given to improving the small sample properties by using bias adjusted estimates and multivariate t-distributions.

Recommendations

- Fallback tests for co-primary endpoints uniformly improve the classic test and hence should be applied whenever testing hypotheses on co-primary endpoints.
- When testing hypotheses on multiple binary endpoints with small sample sizes, optimal exact tests should be used.
- The simultaneous inference approach for multiple marginal general estimation equations models should be considered when testing hypotheses on multiple endpoints with complex data structure. Small sample adjustments for this approach should be included for inference in small trials.

Adaptive designs in drug development for rare diseases

Sequential and adaptive clinical trial design can both enhance efficiency as well as increase the probability of success. We developed novel methods for the design and analysis of clinical trials in small populations, with the aim to make the most efficient use of available information.

Group sequential trial designs allow for test decisions at interim analyses, such that a study may be stopped early, either if a treatment effect can be demonstrated based on interim data, or if the interim data suggests that the treatment is not effective. This may result in lower expected sample sizes and shorter expected trial durations. Adaptive trial designs allow in addition for modifications of aspects of the trial design, such as the sample size, the follow-up period or the recruitment rate based on an interim analysis. In Nikolakopoulos et al., (2016), we evaluated the operational characteristics of sequential designs in the setting of very small to moderate sample sizes with normally distributed outcomes and demonstrate the necessity of simple corrections of the critical boundaries. In addition, the group sequential designs were optimized given a maximum sample size and preliminary information on the treatment effect. In van Eijk et al. (2017), the potential timesaving properties of sequential designs for confirmatory trials was exemplified for clinical trials in amyotrophic lateral sclerosis (ALS), a serious, progressive and lethal condition for which no proven treatment

exists. It is shown that sequential designs are potentially beneficial for neurodegenerative disorders more generally (Figure 4). Reducing the trial length has not only important financial consequences, but more importantly can make clinical trials more ethical by reducing the patients' exposure to non-effective treatments or limiting their time on placebo.

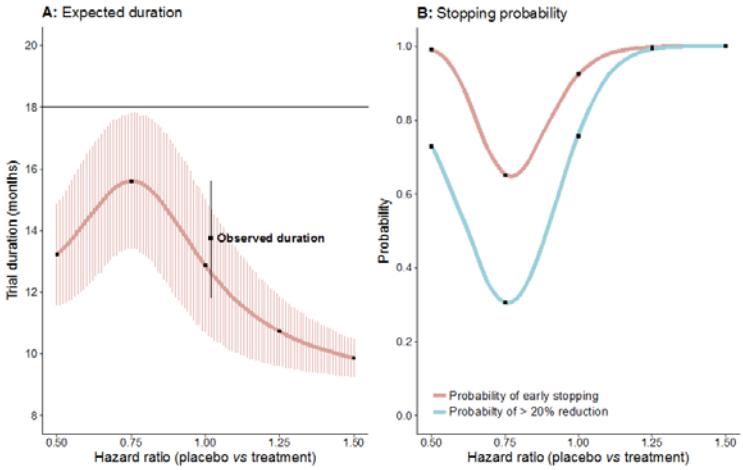


Figure 4. Improvement in ALS with sequential design. Survival data was simulated (10,000 simulations per scenario) based on a Weibull model fitted to the EMPOWER placebo group, with an identical sample size (n = 942) and accrual rate (6 months), using a K = 4 GSD with interim analyses set at 40%, 60%, 80% and 100%. Maximum trial duration was set to 18 months. We scaled the observed duration by multiplying the reductions from table 2 with 18 months. GSDs are least effective when there is a small treatment effect (HR of 0.75) and most effective when there is large positive or negative treatment effect. In none of the scenarios could the sample size be reduced.

In Urach and Posch (2016), group sequential designs for multi-arm trials with different stopping rules were investigated. Multi-arm trials need smaller patient numbers compared to multiple two-arm studies, as the former share the control group. For rare diseases they may provide a very efficient approach if multiple treatments, possibly even from different companies, are investigated in the same disease. The required sample size may be further reduced if interim analyses following certain stopping rules are introduced. We considered stopping rules as simultaneous and separate stopping (see Figure 5) that match the trial objectives to either show a treatment effect for at least one or for all effective treatments. We derive critical boundaries tailored to specific stopping rules that account for the resulting multiple testing problem. For example, under the simultaneous stopping rule, where the trial is stopped as soon as one treatment is demonstrated to be effective, an improved optimal multiplicity adjustment has been derived. Similarly, for the test of multiple endpoints in a group sequential trial, more powerful testing procedures were derived by accounting for the stopping rule (Urach and Posch, 2017).

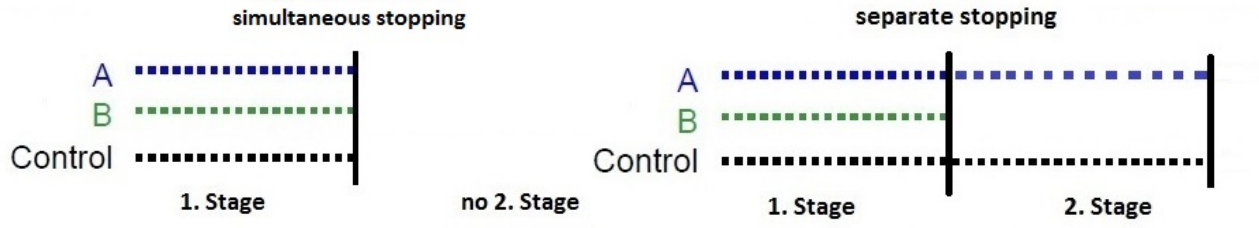


Figure 5. Illustration of the separate and simultaneous stopping rule in a group sequential trial with two experimental treatment arms and a control. With the simultaneous stopping rule, the trial stops as soon as one treatment is shown to be effective. With the separate stopping rule, the trial stops at an interim analysis, if for all treatment arms an effect has been demonstrated. (Urach and Posch, 2017).

Adaptive testing procedures allow for additional flexibility compared to group sequential designs and are more robust with respect to planning assumptions which are often based on

very limited information. However, this flexibility comes with increased complexity in the construction of hypothesis tests for confirmatory trials. In Magirr et al. (2015), we developed a novel adaptive test for time-to event endpoints and compared it to several alternative proposals. We showed that when choosing a testing approach, one needs to compromise on either the type I error rate, the power, the utilization of all data for the interim decision, or inclusion of all events in the final test. In joint work with the Inspire project, Joergens et al. (2017), derived an adaptive test for survival data that allows one to use all interim information to adapt the trial, and in addition for an interim test of the null hypothesis, as in a group sequential trial. Statistical models for the joint distribution of surrogate and clinical primary endpoints are proposed. These enable adaptive designs that use information from early observable surrogate endpoints for adaptation decisions in interim analyses.

Furthermore, the robustness of adaptive designs in small clinical trials with respect to the distributional assumptions (e.g., normally distributed endpoints) was demonstrated in a simulation study (Martin Posch, Gernot Wassmer, Short Course on Adaptive Designs in Small Populations, presented at the Inspire Project Conference, 2017).

A further application of adaptive designs are adaptive dose finding studies, where we contributed to a workshop to assess regulatory aspects of their use (Musuamba et al., 2017).

Recommendations

- For group sequential tests in very small samples, the t-correction is a valid heuristic method for controlling type I error rates.
- For a given maximum sample size, the boundaries of a group sequential design can be optimized, given prior information and taking into account the t-correction.
- Group sequential designs with futility boundaries can save time and resources in the field of ALS, where most trials fail, as was shown by re-analysing trial data.
- Group sequential designs can be recommended for specific cases where recruitment is faster than outcome measurement. In an ALS trial, where all patients have been recruited but multiple outcome measurements are performed at long term follow-up, group sequential tests effectively reduced the number of follow-up visits.
- Stopping rules of multi-armed group sequential trials that take into account specific alternatives can maximize statistical power.
- The choice between analysis methods for adaptive designs with survival endpoints should depend on the desired characteristics, such as type I error rate control, power, and utilization of all data for interim decisions. Improved testing procedures allowing the use of all interim information for adaptation decisions and interim testing of the null hypothesis should be considered for trials with time to event endpoints.
- The use of information from surrogate endpoints should be considered to make most efficient use of the interim data for decisions on design adaptations. Simulation studies as proposed should be used to evaluate adaptation strategies in the interim analysis.
- Regulatory agencies welcome proposals for sound innovative trial designs. Model based and adaptive trial design should be considered for dose-finding studies to improve dose selection instead of traditional pair-wise comparison designs.

Clinical trial design for heterogeneous populations

A disorder related to a single pathophysiological (genetic, metabolic, signalling) pathway may lead to diverse conditions, presenting a heterogeneous patient population. The number of patients sharing the same manifestation of the underlying disorder may be small, limiting the power of a clinical trial with inclusion criteria based on the disease phenotype. In a *basket*

trial for rare diseases, patients with different rare conditions, which share a common pathophysiological pathway, are pooled to increase the number of patients eligible for a trial as well as the number of patients that may benefit from a positive trial result. This approach was exemplified in a trial evaluating the safety and tolerability of a monoclonal antibody treatment for four rare complement-mediated disorders (Derhaschnig et al., 2016).

Recommendation

- Patients with different rare conditions which share a common pathophysiological pathway, should be pooled in a basket trial to increase the number of patients eligible for a trial as well as the number of patients that may benefit from a positive trial result.

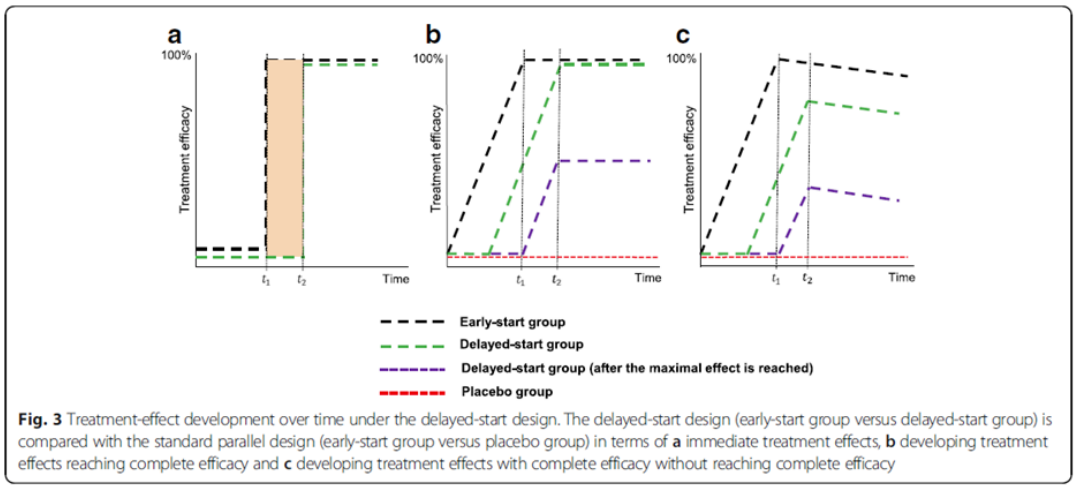


Fig. 3 Treatment-effect development over time under the delayed-start design. The delayed-start design (early-start group versus delayed-start group) is compared with the standard parallel design (early-start group versus placebo group) in terms of **a** immediate treatment effects, **b** developing treatment effects reaching complete efficacy and **c** developing treatment effects with complete efficacy without reaching complete efficacy

Figure 6. Graphical illustration consequences of delayed-start design.

The delayed-start design

Spineli et al. (2017) evaluated whether there are real advantages of the delayed-start design, particularly in terms of overall efficacy and sample size requirements, as a proposed alternative to the standard parallel group randomised clinical trial in the field of rare diseases. We clarify that delayed-start design is not optimal for drugs that establish an immediate treatment effect, but for drugs with effects developing over time, instead. In addition, the sample size will increase if the design includes a reduced time on placebo (which is the true advantage of the delayed start design), as a consequence of the smaller treatment effect (Figure 6).

Recommendation

Deviations from the classical parallel group trial, such as the delayed start design, may be of advantage in specific cases, but are inferior to the parallel group design if the time on placebo is (substantially) reduced.

1.1.3.3 Series of trials methodology

The advances in (sequential) meta-analysis

It is well acknowledged that to overcome the difficulty to conduct large randomized clinical trials for new treatments in rare diseases, adjusted strategies may be acceptable. Yet, the clinical drug development programs for rare diseases as submitted to regulatory authorities usually do include more than one clinical trial to support the benefit risk assessment and ultimately a licensing decision. The ASTERIX project has therefore dedicated research to understanding and improving the role of (sequential) meta-analyses in rare disease clinical research. The original expectation was that, particularly due to the limited size of individual trials, meta-analysis across trials could (1) further strengthen the evidence base and (2)

improve drug development programs when applied in a prospectively planned sequential fashion. The research of ASTERIX uncovered more fundamental issues associated with meta-analysis in the sparse setting of rare diseases. Based on the results, guidance for meta-analysis for small numbers of small trials can be given. In addition, the fundamental quality of simulation studies that are used to study the properties of meta-analyses can be improved. Finally, it has become clear that the guideline on clinical trials in small populations (CHMP, 2006) needs adaptation with respect to meta-analysis.

Meta-analysis in drug development for rare diseases

Meta-analysis is usually performed on top of a series of clinical trials, each of which is adequately powered to support efficacy. For the assessment of safety parameters – especially for rare events – integration of data across studies is expected. Present guidance (albeit dating from 2001) is reserved regarding meta-analysis serving as pivotal evidence, but does indicate the following six important regulatory purposes (*cited* from CHMP (2001)):

1. To provide a more precise estimate of the treatment effect.
2. To evaluate whether overall positive effects are also seen in pre-specified subgroups of patients.
3. To evaluate an additional efficacy endpoint that requires more power than an individual trial can provide.
4. To evaluate safety in a subgroup of patients or a rare event in all patients.
5. To improve the estimation of the dose-response relationship.
6. To evaluate apparently conflicting study results.

The Guideline on clinical trials in small populations (CHMP, 2006), seemingly in contrast, puts a well conducted meta-analysis of clinical trials at the top of the hierarchy of evidence, suggesting that a well-planned and well-conducted meta-analysis provides “even stronger” evidence as compared to (individual) controlled trials. Against this background, the strength of evidence from clinical trials for new drugs in rare diseases may at the very least benefit from (routine) prospective planning of a meta-analysis in their development program, where *sequential analysis* may further help to control the level of evidence (at the program level) associated with the meta-analytical strategy, as well as increase efficiency. In this respect it is useful to delineate the difference between sequential meta-analysis and adaptive trial design. In an adaptive trial in essence one trial is designed and executed, which allows pre-planned and well specified changes in design features during the trial. Planned changes in design features should preferably be prospectively specified, such that operational characteristics of the trial can be assessed at the planning stage and to reassure interpretability of results if the overall outcome is significant. *Sequential meta-analysis* includes a well-considered decision to conduct separate independent trials, allowing differences in design as well as types of in-between decision making not accommodated in adaptive designs.

Before solid recommendations could be given, it became clear from extensive reviews of meta-analysis for rare diseases that crucially relevant aspects for their use in small populations need to be considered (e.g., Pateras at al., 2017). These include between-study heterogeneity in case of a small number of studies, the small size of individual trials, and matters of extrapolation (if the time order of the studies is taken into account).

Conceptual framework for application in clinical development

We distinguish the following three classes of application:

A. Meta-analysis is employed to integrate the evidence after the trials are completed (whether prospectively planned or not). The meta-analysis (or analyses) are used to evaluate

the robustness of results and/or potentially answer additional questions, but not usually to provide primary evidence of efficacy.

B. Prospectively planned meta-analysis are conducted at the end of the trial program as part of – or even as sole basis of – providing confirmatory proof of efficacy.

C. The program of clinical trials for the rare disease/small population is prospectively planned, with sequential meta-analysis to both allow early stopping of the program (after a study) as well as basis to provide primary evidence of efficacy.

Across the three classes from A-C there would typically be: increasing concern about demonstrating control of the type I error rate, increasing concern about the compatibility of the different trials (a matter of extrapolation between trials that may have different populations, different objectives and/or different designs) and increasing concern about the effect of between-trial heterogeneity (of similar trials). We could aim for methods that could be demonstrated to be sufficiently robust or sufficiently precise in a broad range of situations with respect to these challenges. This would at least allow (sequential) meta-analysis to be an option for strengthening the evidence in drug development for rare diseases. In all of this, the advantage in drug development is that individual patient data is available for analysis, allowing more in-depth meta-analysis and modelling than based on summary statistics.

Current insights based on research and results

Heterogeneity between studies of the (underlying true) treatment effect is arguably the greatest concern for meta-analysis in our setting. For small populations this concern is expected to be aggravated, since there would typically be only a few studies available: it is unlikely that the heterogeneity (parameter) can be reliably estimated from the data. This may, however, depend on the estimator of heterogeneity employed. This problem was tackled from different perspectives, leading to the following results.

(i) In case of heterogeneity in a meta-analysis of two studies, classical fixed effect analysis will increase the type I error – to unacceptable levels. The Hartung and Knapp approach does control the type 1 error irrespective of heterogeneity, but at the expensive of very low power. Hence, it does not present a reasonable alternative yet (Gonnermann et al., 2015; Figure 7).

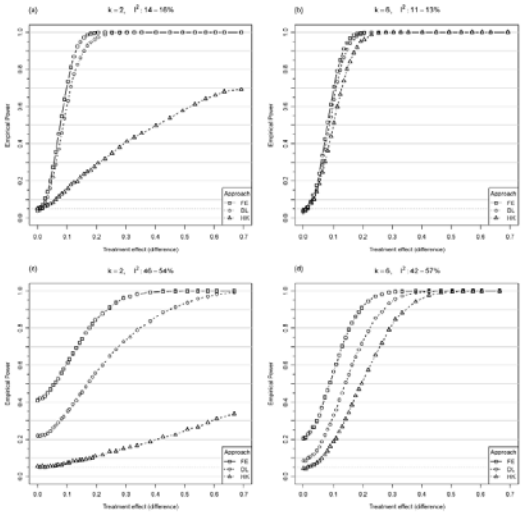


Figure 1. (a–d): Influence of heterogeneity in meta-analysis with two and six studies on empirical power. FE, fixed effects approach; DL, DerSimonian and Laird approach; HK, Hartung and Knapp approach. In the left column, simulation results with two studies are presented, whereas in the right column, situations with six studies are investigated. No heterogeneity is assumed in the top row, and in the bottom row, the impact of moderate heterogeneity is shown.

Figure 7. Influence of heterogeneity with 2 and 6 studies on power.

(ii) In case random effects analysis is applied to a small series of small trials, there are substantial differences between the heterogeneity estimators with respect to type 1 error control, bias of the treatment effect estimate and coverage. For both dichotomous and continuous outcomes the two-step DerSimonian and Laird estimator and the Paule-Mandel estimator can be recommended; for continuous outcomes also the REML estimator is an option. It should be noted, however, that for truly small series of trials also these estimators are subject to bias (Pateras et al., 2017). For the sparse event setting Rukhin's positive Bayes estimator, the improved Paule-Mandel, the Sidik-Jonkman and Hartung-Makabi estimation methods provide potential alternatives. They do not strictly control the type 1 error rate, but bias and type 1 error rate remain within reasonable limits across a range of levels of heterogeneity and likelihoods of observing 0 events in any treatment arm.

(iii) To arrive at more robust interval estimates of the treatment effect based on a meta-analysis, Bayesian analysis was investigated, evaluating a range of possible (informative) priors (Pateras et al., 2017). The heterogeneity problem cannot be completely solved, as expected. In sparse settings the use of a Bayesian binomial-normal meta-analysis with a limited informative Uniform (−10, 10) prior on log(τ^2) provides *robust results*, especially when operational characteristics (type 1 error and power) have to be preserved.

The challenge of heterogeneity in meta-analysis (or similarly between strata in an individual trial) also has an underlying conceptual problem, as alluded to by Senn (3). The strict null hypothesis of no treatment effect, i.e. not on average across the population nor in any subgroup of patients, does in fact not allow heterogeneity in treatment effect (between strata or studies). Hence, the absence of heterogeneity is an essential part of the null hypothesis that is being tested. This leads to considering hypothesis testing and (interval) estimation somewhat differently: Estimation of the treatment effect if it exists has to consider the (potential) heterogeneity as this is part of the uncertainty and variance in treatment effect across the population. But the null hypothesis of no treatment effect could (in the strict sense) be tested in a fixed effects model, as under the strict null of total absence of treatment fixed effect type 1 errors reflect the error under this model. Within ASTERIX we therefore looked at the meta-analysis challenge both from a hypothesis testing and estimation perspective, and in essence treated heterogeneity as a nuisance parameter (whose direct estimation is of lesser interest).

A second consideration that emerged in the ASTERIX research, is that operational characteristics of (random effects) meta-analysis models are often established based on simulation studies. Through our work it appeared that different researchers actually used different data generating models, which do not differ very much for large sample sizes but do differ for small numbers of small studies (Pateras et al., 2017). Proper choice of the data generating model and proper description is essential to assess comparative performance of statistical methods in this field.

Recommendations

- Meta-analysis has a place in small populations research and should follow the well-known recommendations for meta-analysis in general, such as prospectively planned, as this will also improve individual study design.
- Although the impact of heterogeneity precludes providing general recommendations on methods to achieve optimal evidence from meta-analysis, selected Bayesian (informative) priors provide some level of robustness and can be recommended.

- The impact of heterogeneity is not limited to small numbers of small trials; similar problems arise within large trials with few strata. This result is surprising and counter-intuitive, because in the presence of limited patient resources methodologists usually would advise limiting the number of stratifying variables and the number of strata.
- The current small populations guidance should be revised, as it places a well conducted meta-analysis at the top of the hierarchy in terms of strength of evidence. This will need considerable nuance and needs to reflect the different uses of meta-analyses in drug development identified: (a) retrospective, (b) prospectively planned and (c) with a sequential nature, where design of studies depends on previous studies.

Including historical data and combining data across drug development phases

Bayesian (sequential) meta-analysis could be viewed as an approach to systematically include historical data in the prospective design and analysis of a new trial. ASTERIX also looked at this potential approach at a more fundamental level. Motivated by the challenge that in the case of small populations available data is scarce and heterogeneity cannot easily be estimated or understood, we searched for alternative ways to borrow evidence from historical studies. The two key thoughts were to develop a prospective approach that (1) would automatically down weight historical data, in case of a "large difference" with results for the new study (that have yet to be collected) and (2) would enable preserving the type 1 error of the new trial to be conducted. The concept of *power priors* was considered to be particularly useful for borrowing evidence from a single historical study. Power priors employ a parameter $\gamma \in [0,1]$ that quantifies the heterogeneity between the historical study and the new study. However, the possibility of borrowing data from a historical trial will usually be associated with an inflation of the type I error. We developed a new, simple method of estimating the power parameter suitable for the case when one historical dataset is available (Nikolakopoulos et al., 2017). The method is based on predictive distributions and parameterized in such a way that the type I error can be controlled by calibrating to the degree of similarity between the new and historical data. The method is demonstrated for normal responses in a one or two group setting. The key results are that (a) type 1 error can indeed be controlled, (b) the method is indeed more efficient compared to reasonable alternatives (such as "test heterogeneity and then pool" approaches, Figure 8). Importantly, if the new trial is to be comparative (e.g., new drug versus placebo), historical data needs to be comparative as well. Although this seems restrictive, recent research has shown that including only historical "control arm data" in design and analysis of a new trial has only very limited benefit (Galwey, 2016), but true benefit can arise from historical treatment effect estimates from properly controlled trials.

Generalization to other models appears straightforward, which actually may make this approach more fundamental to achieving efficiencies by including historical data in small populations research. An example includes sample size re-estimation in small population trials. The sample size of a randomised controlled trial is typically chosen in order for frequentist operational characteristics to be retained. For normally distributed outcomes, an assumption for the variance needs to be made which is usually based on limited prior information. Especially in the case of small populations the prior information might consist of only one small pilot study. As part of ASTERIX we suggested sequential methodology to deal with this problem (Nikolakopoulos et al., 2016). Alternatively, the idea of power priors introduced above was extended to this situation (Brakenhoff et al., 2017), providing a novel way to re—estimate the sample size and increase the probability of success for individual small population clinical trials.

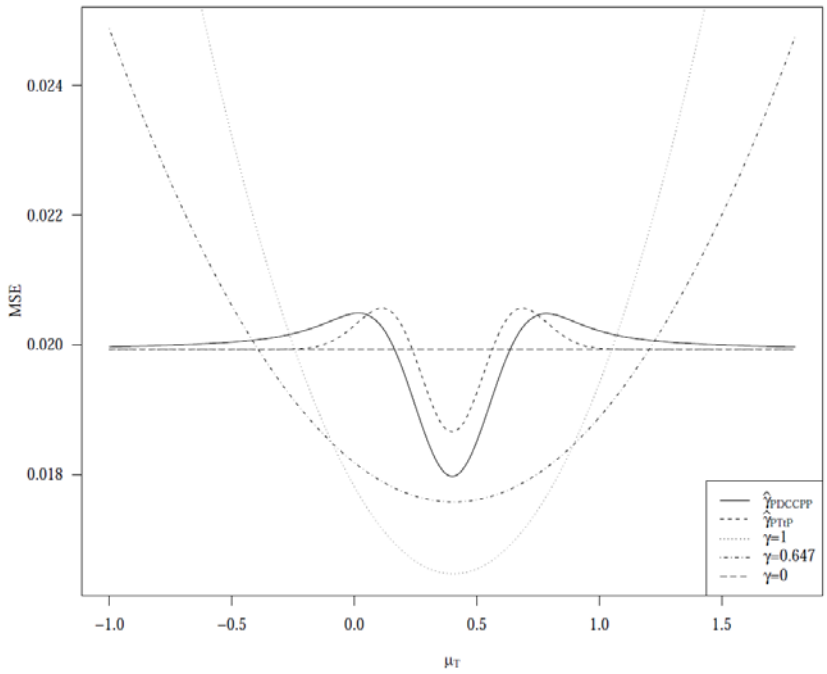


Figure 8. Performance comparison of prior data conflict calibrated power prior (pdccpp). Compared to other approaches in terms of the mean squared error (MSE) of estimating the treatment effect (as function of the true effect). Alternatives are no pooling, power priors with fixed weighing and test then pool.

The power prior approach above could likewise be extended to combining data from Phase II and Phase III trials. However, given that the decision to go through Phase III actually depends on Phase II data, more tailored approaches may be needed, such as proposed by Stallard and Todd (2010) or Kunz et al. (2014). Also, previous work of our group explored various directions, where Phase II biomarker results are used to increase efficiency of Phase III trials (Nikolakopoulos, 2013). Important outcomes are not always observed in early randomized clinical trials because they can be either expensive to gather or observed after a long-term follow-up period. On such occasions short-term or less expensive outcomes can be used to provide supplementary treatment efficacy evidence on the long-term outcome. However, when the confirmatory (phase III) trial is based on significant result of the earlier trial (Phase II) short-term outcomes, final inferences risk producing large error rates. In this context, we propose an approach that controls the conditional error rates to detect a treatment effect. The method appears to (1) allow the use of short-term outcomes only when really needed (small evidence from phase II), (2) increase empirical power and control type I error relatively better than the simple double-regression (3) reduce the problem with bias in the estimation of the variance and the treatment effect of the double-regression approach. We illustrate the methodologies through an orphan drug example for Fabry disease. This research is still ongoing and will be completed in the first quarter of 2018.

Recommendations

- Prospective rules for the inclusion of phase II data in the design and analysis of a phase III clinical trial can mitigate the concerns related to bias. The concept of dynamic borrowing can be utilized. Thus, Phase II data can be included in the design phase of a Phase III trial so as to optimise the (sequential) design, without hampering its operational characteristics nature.

- The concept of dynamic borrowing developed has also been implemented in the case of adaptive designs with sample size re-estimation, to robustly estimate feasibility of a small trial based on interim results.
- Nuisance parameters, such as the variance, are also subject to a lack of robustness when phase II data are included in the design of a phase III trial. This is of interest as such considerations are usually overlooked.

1.1.3.4 Patient perspective

The importance of including the patient perspective in all aspects of clinical research has been widely acknowledged. Unfortunately, this does not mean it has been equally widely implemented effectively. At the start of ASTERIX we considered that the focus on rare diseases would be a fertile ground for expanding involvement of patients in clinical trial design and conduct. Thus, we established a Patient Think Tank advisory to all of the research conducted, and included key topics specifically addressing the patient perspective. These were:

- Optimal use of rare disease registries to enhance the efficiency and feasibility of clinical trials in small populations for regulatory decisions.
- Development of a practical model to involve patients and patient representatives in the choice and prioritisation of outcomes in rare disease clinical trials.
- Determination of the added value of the patient centred outcome instrument ‘Goal Attainment Scaling’.
- Elicitation of opinions of patients and patient representatives about what aspects of design of clinical trials are most relevant to them.

In our deliberations we also concluded that ethical aspects are closely related to both methodology as well as patient perspective, and may have specific considerations for rare diseases. Finally, the direct involvement of patient representatives in methodological research was new to all, which stimulated an evaluation from which we could learn for future projects.

Optimal use of patient registries to inform design

Low prevalence, lack of knowledge about the disease course, and phenotype heterogeneity hamper the development of drugs for rare diseases. Rare disease registries can be helpful in understanding the course of the disease, and providing information necessary for clinical trial design. To develop recommendations for coordinators for design and management of rare disease patient registries we performed a scoping review of the literature. In addition, several meetings were held with experts from the ASTERIX consortium and structured interviews were conducted with coordinators and academic experts of selected rare disease patient registries. A list of recommendations was drafted, as well as a checklist with all types of information that should be included in a rare disease registry to make the information useful for future trial design (Jansen-van der Weide et al., 2017).

Registries have been used, and are being used successfully for regulatory purposes. Notably for post-authorisation commitments to continue research into safety of new drugs, but also in exceptional cases as part of market authorisation applications. In these exceptional cases registry data, either collected retrospectively or prospectively, were used to replace the use of placebo or active comparators. However, it is at present not clear in what circumstances this might or might not be acceptable.

During our investigations, difficulties in setting up a sustainable long term registry became clear. Very often it depends on temporary, research project driven funding. This

tends to limit the scope and amount of data collected, and creates uncertainty in keeping the registry up to date.

Our scoping review and expert evaluation led to the following key findings.

- A rare disease registry can be instrumental in trial design by: informing the sample size calculation, increasing efficiency as a data collection tool in clinical trials, replacing the placebo or active comparator group with a historical control group, or serving as key data collection tool in post marketing surveillance.
- To enable optimal use of a rare disease registry longitudinal data collection is indispensable, and specific data collection, prepared for repeated measurements, is needed. The developed checklist can help to define the appropriate variables to include.
- Disease-specific, patient centred, rare disease registries are preferred over product-specific registries. In a disease-specific registry all consenting patients with the disease are included, and not only the patients who receive a certain treatment.
- Valid measurement instruments should be used, and measurements, data collection and data management should make use of global data standards to optimise comparability and interoperability with clinical trial databases.

Some future challenges remain, that are becoming increasingly important. A more solid understanding of situations and uses of registry data that may be acceptable as part of marketing authorisation applications would be helpful, both for industry as well as registry builders. To maximize efficiency, post-marketing safety assessment should be conducted by means of (often already existing) disease-specific registries instead of product-specific registries. This would allow for better comparisons and protect the evaluation process from commercial influences. It can also enhance possibilities for gathering information on the use of the products under special circumstances (such as extended licensing) or even for collecting clinical trial data. Health authorities, pharmaceutical companies, researchers and patients should aim to make this a common accomplishment.

Recommended methodology to elicit patient preferences and outcome weighing

In drug trials, it is relevant to ask patients and/or their caregivers which aspects concerning their disease/condition they consider important to evaluate when a new medication is being investigated. This input can lead to the use of patient-centred outcome measures in a trial, or clear preference weighing of desired improvements. In rare diseases the choice of outcome measures may be even more important, due to the combined challenge of small sample sizes and heterogeneity of the disease course. Based on 1) models for involvement of patients and patient representatives in medical research found in the literature, 2) an iterative process of extensive communication with the members of the ASTERIX Patient Think Tank, and 3) a road-testing pilot in a focus group consisting of patients with Spinal Muscular Atrophy we developed a method to involve patients in the determination and weighing of outcome measures in the design stage of a phase II or phase III trial (Gaasterland et al., 2016a). The model, called POWER model (an acronym of Patient participation in Outcome measure WEighing for Rare diseases) consists of three steps: 1) Preparation, 2) Consultation of patients in two phases, 3) Follow-up. The first road-test in a small focus group was evaluated positively both by patients and by clinical researchers. We advise replication of testing and further refining the model for use in early stages of the planning of future trials, preferably before the trial protocol is finalized. It can provide a valuable tool for use in clinical research, especially in rare diseases, to include patients in the design stage of trials in a more structural way. The POWER model can be used to develop a collaboration in which patient representatives together with other stakeholders can brainstorm on outcome measures and choose or prioritize them, when there are no standard sets of outcomes to choose from. We think that this model can be a useful tool for participation, *changing the direction from research in patients towards research with patients.*

The value of Goal Attainment Scaling in rare disease clinical trials

In heterogeneous populations, there may not be a single clinical endpoint that is suitable to assess the effect of a treatment in all patients. Different disease stages or diverse disease manifestations may require different outcome measures. Duchenne's muscular dystrophy is a case in point: the broadly used 6 minute walking test can only be assessed in boys that are still able to walk, but boys who are already confined to a wheelchair may still benefit considerably from treatment. If individual outcomes could be defined for each patient, they could all be included in the same trial. Goal Attainment Scaling is an assessment framework in which individual treatment goals are defined for each patient and the attainment of each goal is assessed on a common discrete scale. Thus, it is truly patient centred. Goal Attainment Scaling contains a variable number of self-defined goals and for each a very explicit description of a fixed number of standardized possible levels of goal attainment that are formulated before the intervention, usually in consultation between the patient and the clinician. For each goal the expected level of goal attainment in case the treatment is effective, and at least two other levels need to be described in such a specific way that an independent observer can assess the outcome.

Goal Attainment Scaling has been proposed as an instrument in rare disease drug trials if there is no disease-specific measurement instrument available that is responsive enough in small heterogeneous groups. However, neither statistical analysis properties have been well investigated, nor has it been properly validated as appeared from our review (Gaasterland et al., 2016b). The concept was received with great enthusiasm by researchers as well as patients: it may address an important unmet need for clinical research in heterogeneous, progressive diseases.

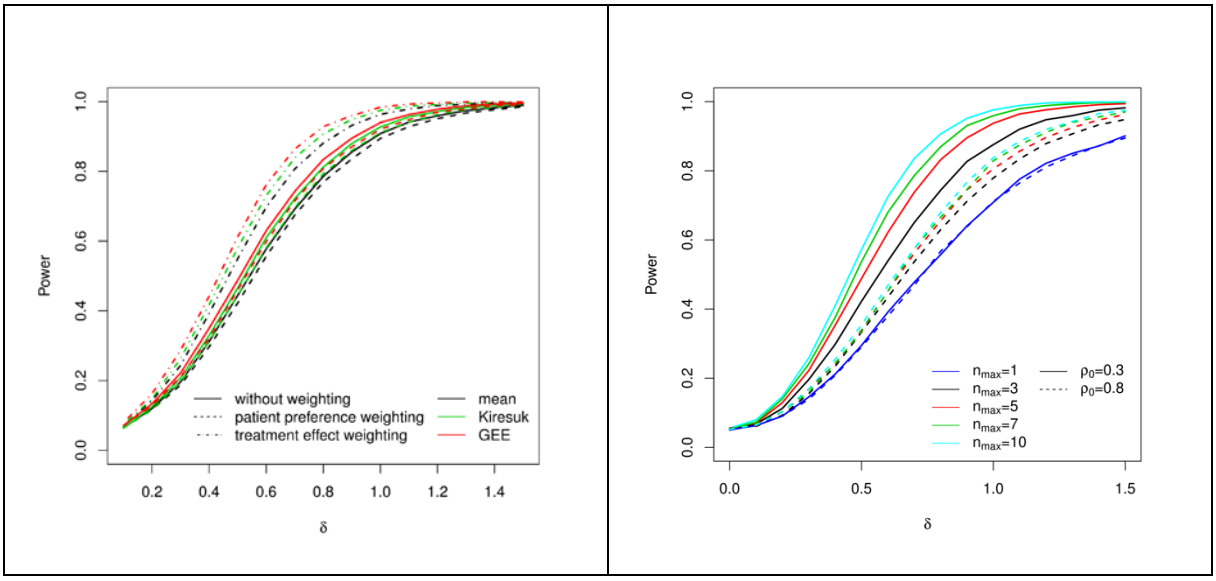


Figure 9. The left Figure shows the power of the three considered testing procedures without weighting, with patient preference weighting and with treatment effect weighting. The right figure gives the power of the GEE approach for different maximum number of goals and a correlation between the goals of 0.2 and 0.8.

Urach et al. (2017) provided solutions to the statistical analysis challenges of Goal Attainment Scaling in randomized controlled trials. To avoid bias, the goals, the criteria for the assessment of their attainment, and the weights are assumed to be chosen before the patient is randomized to one of the treatment groups. Double-blind trials are preferred, but as a minimum requirement a trial with a Goal Attainment Scaling endpoint should be observer-

blinded. Hypothesis tests for the two-group comparison are based on aggregating the observed goal attainment levels for each subject and weighting subjects depending on the number of goals and the correlation between goals. Further, patient preference weights may be included to address the patient-perceived importance of each goal (Figure 9). In principle, an increasing number of goals per patient conveys more information. Consequently, a small number of weakly correlated goals per patient, avoiding goals that are unresponsive to the treatment, may improve the power in spite of a small sample size. The power may be further improved by carefully choosing the actual goal attainment scale such that the expected attainment level under control is closer to the lower end of the scale and the expected level under an effective treatment is closer to the upper end of the scale.

While the Goal Attainment Scaling endpoint allows one to include a very heterogeneous patient population in a clinical trial, this comes at the cost of a more complex interpretation of the trial results, in particular an observed treatment effect might not be easily translated functionally or medically.

Recommendations

- Goal Attainment Scaling can have substantial added value in specific medical conditions.
- A small number of weakly correlated goals should be chosen.
- Goals are set with both patient improvement and mode of action of the drug in mind. Goals that are not sensitive to the treatment can substantially harm the power of the test.
- The impact of the goal attainment level definitions on the statistical power should be considered in the planning phase.
- The analyses based on a generalized estimating equations approach appears to be the most powerful testing strategy for GAS endpoints. For very small sample sizes the more robust t-test based approaches for per-subject means should be used.

We finally developed a plan for the validation of Goal Attainment Scaling in drug trials (Gaasterland et al., 2017). This plan contains a blueprint for the evaluation of several clinimetric properties (content and construct validity, intra- and inter-rater reliability and responsiveness) in a randomised clinical trial of an intervention aimed to influence the underlying pathophysiological problem in mitochondrial disorders, as an example. Goal Attainment Scaling is an instrument intended to measure change on an individual level, instead of the status of the patient at specific points in time (typically before and after the intervention). It does not necessarily measure the same construct in every patient. Validation of the instrument needs to take place not only at the individual patient level, but also on trial level, and several validation trials are required before the validity of Goal Attainment Scaling is sufficiently confirmed. Before the validation of Goal Attainment Scaling is performed, it is crucial that all Goal Attainment Scaling procedures and interviews are standardised, so that it is very clear what instrument is evaluated, and if its validity is considered adequate, how it should be applied to make sure that the procedures are identical to the procedures of the validated version.

Methods for patient involvement in design

Initially, our plan was to perform a survey among patient organisations in Europe about the novel designs developed in ASTERIX. On advice of the Patient Think Tank we decided to change our plans. Instead, we have worked on a more qualitative approach of patient involvement. We interviewed the individual members of the PTT about aspects of trial design that were important to them, and summarized their information for use in the whole project. This raised awareness among the researchers of the issues that really matter to patients. Not all issues raised by the patients were within the remit of the ASTERIX project, e.g. issues regarding access to new drugs after initial regulatory approval. As these matters also do hold

methodological challenges, such as the evidence required for reimbursement decisions, these may become part of future work.

Ethical framework

The difficulties in conducting clinical trials of adequate size, specifically for very or ultra-rare diseases, also raise ethical questions. In addition, matters of consent, privacy and potential pressures to participate in research may be different in rare diseases compared to more common diseases. In Hasford and Koch (2017), the case is made that many ethical aspects – such as consent and statistical properties such as type 1 error an power – are also codified by law, and cannot depend on prevalence of a disease. In an extensive study, against the well-established framework of Emmanuel (2008), we reflected more extensively and attempted to address where novel methodology impacts ethical assessment (van der Graaf et al., 2017). This resulted in an ethical framework to assess innovative designs in rare diseases as a reference for the evaluation of different designs as well as for review by ethical committees. The resulting framework contains eight recommendations – at system, ethical committee or individual trial level - that are especially (but not always exclusively) relevant to rare disease clinical research.

1. Form collaborative partnerships (more important within the small community and potential competing treatments being investigated).
2. Enhance the social value of rare diseases research (at individual trial and system level).
3. Maintain similar standards for scientific validity.
4. Promote rare diseases research (at system level).
5. Recognize specific ethical aspects in the risk-benefit assessment.
6. Ensure expertise on rare diseases research in the ethics review.
7. Ensure voluntary informed consent and provide adequate information (for rare disease it can be more difficult to avoid undue pressure).
8. Ensure respect for privacy (which may be more challenging in rare diseases).

Evaluation of interaction with the Patient Think Tank

The collaboration with the Patient Think Tanks was an unique endeavour, with tangible results – such as patient information leaflets and co-authorship of papers. In an evaluation of this process, we have interviewed researchers of the ASTERIX consortium and members of the Patient Think Tank. Overall, the patient involvement in ASTERIX was considered a success by both the ASTERIX researchers and Patient Think Tank members. Lessons learned for future projects include:

- Patient representatives should be involved at the stage of writing of the project plan.
- The project plan should define specific goals and methods for the patient involvement.
- The formal position of the patient group should be well defined.
- Patient input should be acknowledged and reimbursed.

1.1.3.5 Level of evidence

The question on what constitutes an appropriate ‘level of evidence’ to grant a new treatment marketing approval and make it available to patients, is particularly pressing for (very) rare diseases. Not surprisingly, all three FP 7 projects – IDeAl, InSPiRE and ASTERIX – identified this as an important research area (Hilgers, Roes and Stallard, 2016). ASTERIX collaborated

specifically with IDeAl on some of the topics tackled. Key areas of research were: (1) the present approaches and uncertainties in regulatory assessment and reimbursement (2) the importance of randomised evidence (versus single armed trials or historical data) and (3) the potential of extrapolation to increase efficiency of clinical trials. Finally, we also addressed the importance of data sharing.

Present regulatory approaches and uncertainties

In joint work with IDeAl (Hofer et al., 2016), an analysis was performed on a data set of 157 orphan designated medicines with an outcome for marketing authorisation application (MAA) between 2000 and 2013. A randomised clinical trial was performed in most (70/104; 67%) of orphan market authorisation applications with positive outcome. The study also demonstrated that orphan medicines have a lower success rate compared with non-orphan medicines. Further research on regulatory assessment addressed the uncertainties in benefit-risk assessment as part of the approval procedure (Zafiropoulos et al., 2017). Based on retrospective review of European Public Assessment Reports (EPARs) for new oncology products approved since 2011 (n=64, 26 orphan and 38 non-orphan), the following was concluded. Several uncertainties, crucial to the assessment of new pharmaceutical products, remained at the time of market authorisation. The majority of issues was due to insufficient data, which led to the requirement for submission of post-approval data. Orphan status had no major impact, with the only main difference being a higher number of uncertainties driven by unreliable data. EPARs with a lack of randomised clinical trials were correlated with a higher number of uncertainties, driven by safety issues.

New treatments only become truly available to patients if they are not only approved, but also reimbursed. Since the introduction of benefit assessment to support reimbursement decisions there could be the impression that totally distinct methodology and strategies for decision making would apply (Koch et al. (2016). In order to compare both regulatory processes two examples were chosen, where decision making was based on the same pivotal studies in the licensing and reimbursement process. The main conclusion is that strategies in the field of reimbursement are (from a methodological standpoint) until now more liberal than established rules in the field of drug licensing, but apply the same principles. Formal proof of efficacy (as part of market authorisation) preceding benefit assessment can thus be understood as a gatekeeper against principally wrong decision making about benefit and risks of new drugs in full recognition that more considerations are needed. For clinical trials in very rare diseases the opposite situation may occur: in some instances no formal proof of efficacy is possible and benefit/risk assessment needs to integrate findings from different endpoints or different subgroups without some sort of shelter from a preceding formal proof of efficacy.

These explorations thus indicate the importance to have a solid understanding of the standards to assess evidence in exceptional situations, and to understand in what circumstances the evidence can rely on non-randomised evidence.

The importance of randomised evidence and role of non-randomised data

In joint work with IDeAl, Eichler et al., (2016) formulate assessing benefits and risks of (new) treatment in causal inference terms: The central question is: how does the average outcome on the new treatment (*the factual*) compare to the average outcome had patients received no treatment or a different treatment known to be effective (*the counterfactual*)? Randomised controlled trials are the standard for comparing the *factual* with the *counterfactual*. Recent developments necessitate and enable a new way of determining the counterfactual for some new medicines. For select situations, we propose a new framework for evidence generation, which we call “threshold-crossing.” This framework leverages the wealth of information that is becoming available from completed randomised clinical trials and from real world data sources. Relying on formalized procedures, information gleaned from these data is used to estimate the counterfactual, enabling efficacy assessment of new drugs. Both single arm

trials (using external data as counterfactual) early in the clinical development program as well as randomised trials are part of this framework. It may provide a more solid base to include non-randomised data into decision making, without jeopardizing the level of evidence needed. We propose future (research) activities to enable “threshold-crossing” for carefully selected products and indications in which RCTs are not feasible.

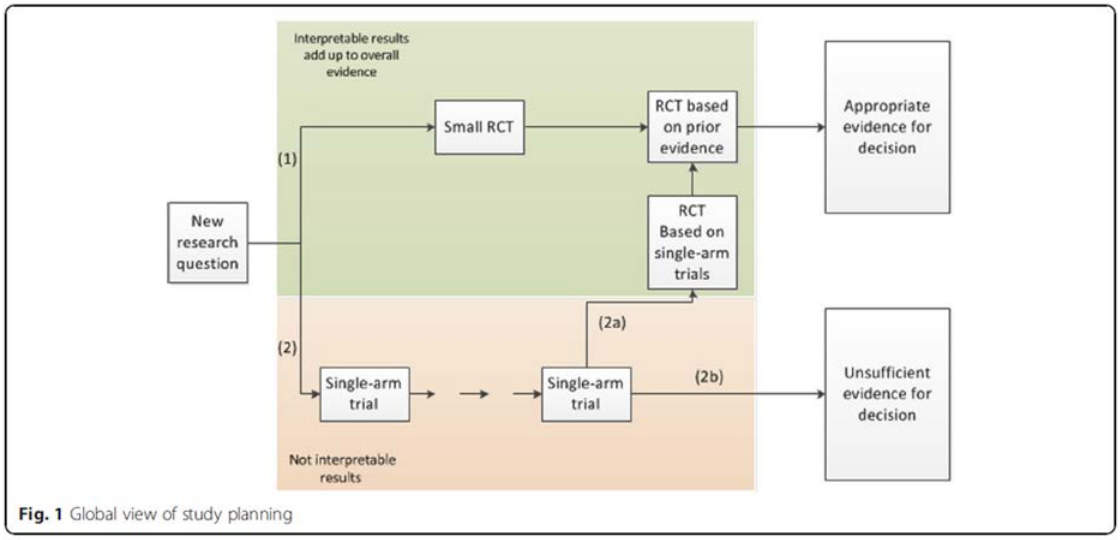


Figure 10. Clinical research strategy with randomised trial building blocks.

Amongst others, this should address overall efficiency of clinical development plans – in terms of total number of patients needed to arrive at a solid evidence base. Lasch et al. (2017) use the example of pioglitazone to resolve leucoplakia and erythroplakia in Fanconi anaemia patients to illustrate that research strategies based on single-arm trials may require more patients from a global perspective than randomised trial based research strategies. Particularly in rare diseases there is merit in planning the research strategy based on RCTs as combinable building blocks that provide unbiased estimates of the treatment effect (Figure 10). Importantly, this strategy helps avoid undocumented selection of patients. Traditional single-arm trials are linked to a wide field of problems, such as debatable interpretation, lower quality of evidence, unrecognized predictive or prognostic factors, and overestimation of the treatment effect derived from a comparison to effects seen outside the current trial. Approaches to alleviate these problems exist (registry based trials, form registries with complete patient population e.g.), that could fit in the “threshold crossing” approach described above.

Experience shows that, also in rare diseases, a randomised clinical trial is only conducted after one or several single-arm trials, if these lead to ambiguous results and the need for a randomised trial becomes inevitable. At the current state of knowledge it is not possible to provide complete account of the conditions under which conditions singe arm trials provide reliable evidence, hence the option to randomise as early as possible should always be considered.

Potential of extrapolation

A full-scale independent drug development programme to demonstrate efficacy may not be ethical or feasible in small populations such as paediatric populations or orphan indications. Various levels of extrapolation from a larger source population to smaller target populations are widely used for supporting decisions in this situation. In joint work with IDeAl, Hlavin et al. (2016) propose clinical trials designs which make use of prior knowledge on efficacy for inference. A framework based on prior beliefs is formulated, in order to quantify the extent to which the significance level for the test of the primary endpoint in confirmatory trials can be relaxed in the target population. It does so by controlling a defined posterior belief in

effectiveness after rejection of the null hypothesis in the corresponding confirmatory statistical test. It is shown that point-priors may be used in the argumentation because under certain constraints, they have favourable limiting properties compared to other types of priors. The crucial quantity to be elicited is the prior belief in the possibility of extrapolation from a larger source population to the target population. It offers an attractive approach to potentially reduce the sample size in paediatric populations, and ideas could be generalized to other settings as well (e.g., re-purposing drugs).

A plea for stepping up the effort to share individual patient level data

One of the most controversial topics discussed among regulatory agencies, the pharmaceutical industry, journal editors, and academia is the sharing of patient-level clinical trial data (Koenig et al., 2015). Broadly sharing data will provide unprecedented opportunities for research and research synthesis. It will also pose new challenges for regulatory authorities, sponsors, scientific journals, and the public. Data sharing also entails intricate statistical questions such as problems of multiplicity. An important issue in this respect is the interpretation of multiple statistical analyses, both prospective and retrospective. Expertise in biostatistics is needed to assess the interpretation of such multiple analyses, for example, in the context of regulatory decision-making by optimizing procedural guidance and sophisticated analysis methods.

Recommendations

- The uncertainties at the time of marketing authorisation for consistent decision making and planning of drug development programs need to be systematically assessed. The regulatory and reimbursement discussion on the level evidence required, with special attention to the importance of randomised trials, needs to be engaged early.
- Randomised clinical trials are associated with less regulatory uncertainties compared to non-randomised clinical studies at the time of marketing authorisation. This is an important indication that decisions to relax the evidentiary standard of randomization in rare diseases should be exceptional. The mere fact that possibly significance cannot be achieved should not preclude avoiding bias in estimation through randomisation.
- Addressing a new research question with randomised trials from the beginning instead of performing uncontrolled studies before planning an randomised trials may reduce the total patient number required to obtain a definite answer.
- In exceptional situations where randomised trials appear not feasible, but where there is reliable evidence on the outcome under relevant control conditions, the concept of “threshold crossing” should be considered as a framework for evidence generation.
- In specific situations, single arm trials may be designed such that selection bias may be reduced sufficiently to render a “threshold crossing” scenario more efficient, as a consequence of more favourable bias-variance trade-off.
- To increase efficiency of paediatric drug development, a quantitative extrapolation framework should be used to specify adjusted significance levels in paediatric investigation plans.

1.1.3.6 Framework, validation and regulatory decision making

Taxonomy of rare diseases

Methodological guidance specific to clinical investigation of a particular disease is an effective method of providing a predictable decision-making framework, and is useful for both developers and regulators. However, currently there is little regulatory guidance on rare

conditions. EMA has issued the general guidance on clinical trials for small populations (CHMP, 2005), which by its nature is not very specific. The huge number of existing rare conditions makes development of specific guidelines for each condition unrealistic. As a part of the ASTERIX project a grouping of medical conditions ("taxonomy") was to be developed, defined by both the medical condition and the indication of the new treatment being investigated. Clusters so defined could guide the applicability of different (novel) methodologies for clinical trials. From a regulatory perspective, this could bridge the gap between general top-level guidance and an unfeasible disease-specific guidance for drug development and regulatory evaluation. The final aim was to provide a framework that facilitates guidance to both researchers and regulators on how to design and analyse clinical trials for rare diseases. Using both quantitative methods as well as consolidation through expert opinion, the clustering as depicted in Figure 5 was developed (Pontes et al., 2017). It consists of six clusters of medical conditions combined with "degree" of rarity of the conditions. Development was based on evaluation of 125 European Public Assessment Reports on orphan medicinal products. All combinations of disease and indication could be uniquely placed in one of the clusters, although diseases might occur in different clusters.

In a recent discussion of experts led by IRDiRC (Jonker et al., 2016) it was agreed that there was a lack of a classification on rare diseases to discuss the potential application of different study methods or designs. The proposed clustering fits as the key cornerstone for this need. In fact, the application of the clustering of medical conditions may have several uses. First, to ease regulatory guidance: almost all new orphan medicinal product developments face a regulatory scenario in which there is a lack of previous references and specific guidance on how to conduct a regulatory development. Second, to ease the validation and acceptance of alternative methodologies suited to small populations, by determining applicability based on cluster characteristics. Grouping by characteristics determining the applicability of methods might theoretically facilitate the validation of new approaches and inferences on the suitability of methods across conditions sharing similar features. Thus, clustering might result in a more affordable and manageable number of scenarios and could help to increase the predictability of regulatory requirements and assessment outcomes in the long term, thus incentivizing investment and innovation.

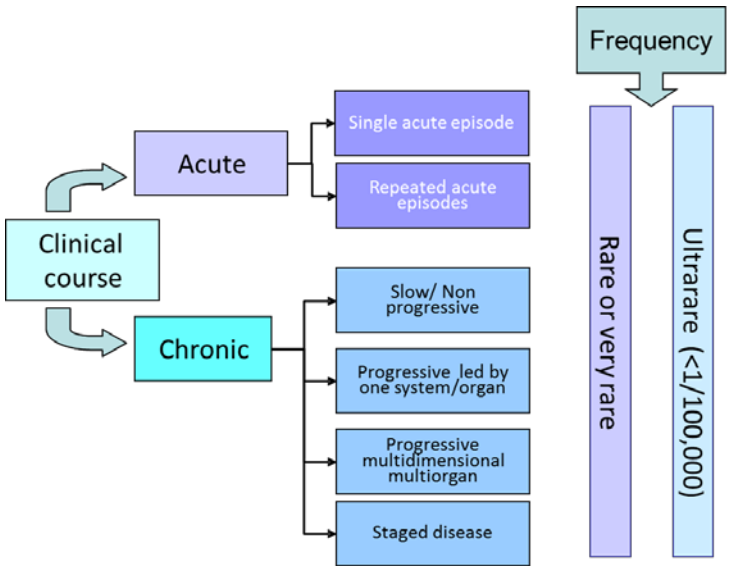


Figure 11. Proposed clustering of medical conditions.

Validation of methods developed by ASTERIX with real life and regulatory data sets

A key outcome of our methodological research is that it is actually applied and helps improve clinical trials to the benefit of patients. Throughout this report, examples are presented: notably in oncology, amyotrophic lateral sclerosis, mitochondrial diseases, cystic fibrosis, amongst others. Most individual methodology papers include reference to concrete applications as well – some based on the actual datasets and prospective trial design. In addition, we set out to evaluate applicability and added value more systematically, using the clustering introduced above as means to structure the evaluation (Oude-Rengerink et al. 2017). The original plan was to gain access to individual patient data of a representative set of clinical data from recent (last five years) filings for market authorisation. Early on in the project it became clear that this would not be feasible within the timelines of the project. Hence, the systematic evaluation of applicability and added value for drug development was based on a thorough evaluation of in tot 26 European Public Assessment Reports (EPARs), representative of the six clusters and different degrees of rarity of the medical condition in Figure 5. In our evaluation, we included all novel methods that were developed within the ASTERIX project and that have been reported in a manuscript by 1 September 2017 for submission to a journal or papers that have already been published. We categorized the methods into four main groups: 'innovative trial designs' (six methods), 'level of evidence' (two methods), 'study endpoints and statistical analysis' (four methods), and 'meta-analysis' (two methods). We selected a representative set of 26 orphan medicinal products approved by EMA between 2001 and 2014. We used the following selection criteria, to assure a minimum level of representation:

- Four to six EPARs for each condition cluster of the six condition clusters.
- For each condition cluster at least one EPAR describing an ultra-rare condition (affecting <1/100,000 persons in the EU).
- At least one re-purposed drug per cluster.

To assess applicability and added value, we basically used a two-step approach. First, evaluation of direct applicability without any material adjustments to the original clinical studies that supported the successful applications. Second, we worked from a more holistic point of view and allowed adjustments to the clinical studies within the context of the objectives of the overall drug development program. For example, a secondary outcome could have been promoted to a primary outcome, if this could increase efficiency and was acceptable within the clinical context and regulatory framework. Results were summarized in a heat map (see Figure 12, for heat map taking the holistic view).

Novel methods for extrapolation, sample size re-assessment, multi-armed trials, optimal sequential design for small sample sizes and Goal Attainment Scaling showed most promise (applicable in more than 25% of EPARs evaluated, even in the restricted assessment). In case the holistic view was taken, optimising application in the full context of drug development, substantial opportunities to improve the development program with novel methods were noted. There were some differences in applicability of the innovative methods between the six condition clusters. Most applicability was seen in the cluster 'chronic: progressive, multiple system/organs', where almost all proposed methods were applicable and add value, while in cluster 'acute: single episodes' the least number of novel methods appeared applicable. For cluster 'acute single episodes' the majority of the methods from 'innovative trial designs' can were applicable. The novel methods developed for 'study endpoints and statistical analysis' methods were most specifically applicable. For cluster 'acute recurrent episodes' all groups of methods can be applicable. For chronic condition clusters, all methods groups can be applicable to some extent, apart from the methods addressed in the 'meta-analysis' group. For the chronic clusters, GAS was found to be particularly applicable, where it would explicitly not be applicable in the other (non-chronic) clusters.

New regulatory models

To further deepen our understanding of added value of new methodology, six EPARs, one representative of each of clusters, were selected to conduct a qualitative assessment of novel methodologies integral across the entire development program (manuscript in preparation). The selection was based on the availability of enough information to identify the key studies done during the clinical development, and the availability of information on the regulatory assessment in the EPAR to support the analysis of uncertainties and weaknesses. The main purpose was to design and evaluate alternative clinical development scenarios. These alternatives would allow multiple novel methods to be introduced, in addition to the basic evaluation that is presented in the heat map. The EPARs investigated were the following treatments and indications: (1) Defitelio® (defibrotide) for in the treatment of Veno-Occlusive Disease post- haematopoietic stem cell transplantation; (2) Ilaris® (canakinumab) for Cryopirine associated periodic syndromes (CAPS); (3) Revestive® (teduglutide) for short bowel syndrome; (4) Soliris® (eculizumab) for the longterm enzyme replacement therapy in patients with a confirmed diagnosis of atypical hemolytic uremic syndrome; (5) Fabrazyme® (agalsidase beta) for the longterm enzyme replacement therapy in patients with a confirmed diagnosis of Fabry disease; and (6) psumit® (macitentan) for pulmonary arterial hypertension. The full account of findings will be reported in an upcoming publication, as this research was of course being completed after all methodological work within ASTERIX was completed. In most of the 6 EPARs studied, 2 or 3 alternative developments scenarios were generated and evaluated.

This allowed both recommendations for use of novel methods, as well as for already existing clinical trial methodologies.

Recommendations

- The six orphan condition clusters could assist medicinal products developers to determine the likely (in-)applicability of novel methods developed within ASTERIX to improve the drug development.
- A revision of the guideline for small populations could incorporate this disease clustering, and systematically evaluate the (non)applicability of all methods available for use in small populations – not limited to methods developed within the ASTERIX project. This will help drug developers to select the optimal method for evaluation of the condition they target, by guiding them to promising methods and also point out which methods are very unlikely to be useful given the characteristics of the condition they target.
- The clustering of medical conditions can provide guidance which methods are not applicable, but seems insufficiently specific to guide clear preference for certain methods.
- Achieving optimal added value from application of novel methods requires a reconsideration of the drug development program in its full context, rather than only for a specific trial.
- Not all clinical development challenges reported in EPARs where covered by the novel methods developed within the ASTERIX project. Further research into methods to optimise these challenges is needed to improve drug development to ultimately be able to efficiently discover effective and safe treatments for all patients with a rare medical condition.
- The unavailability of information on failed or rejected applications hampers methodological research aimed at improvement of the clinical drug development process. More transparency is needed.

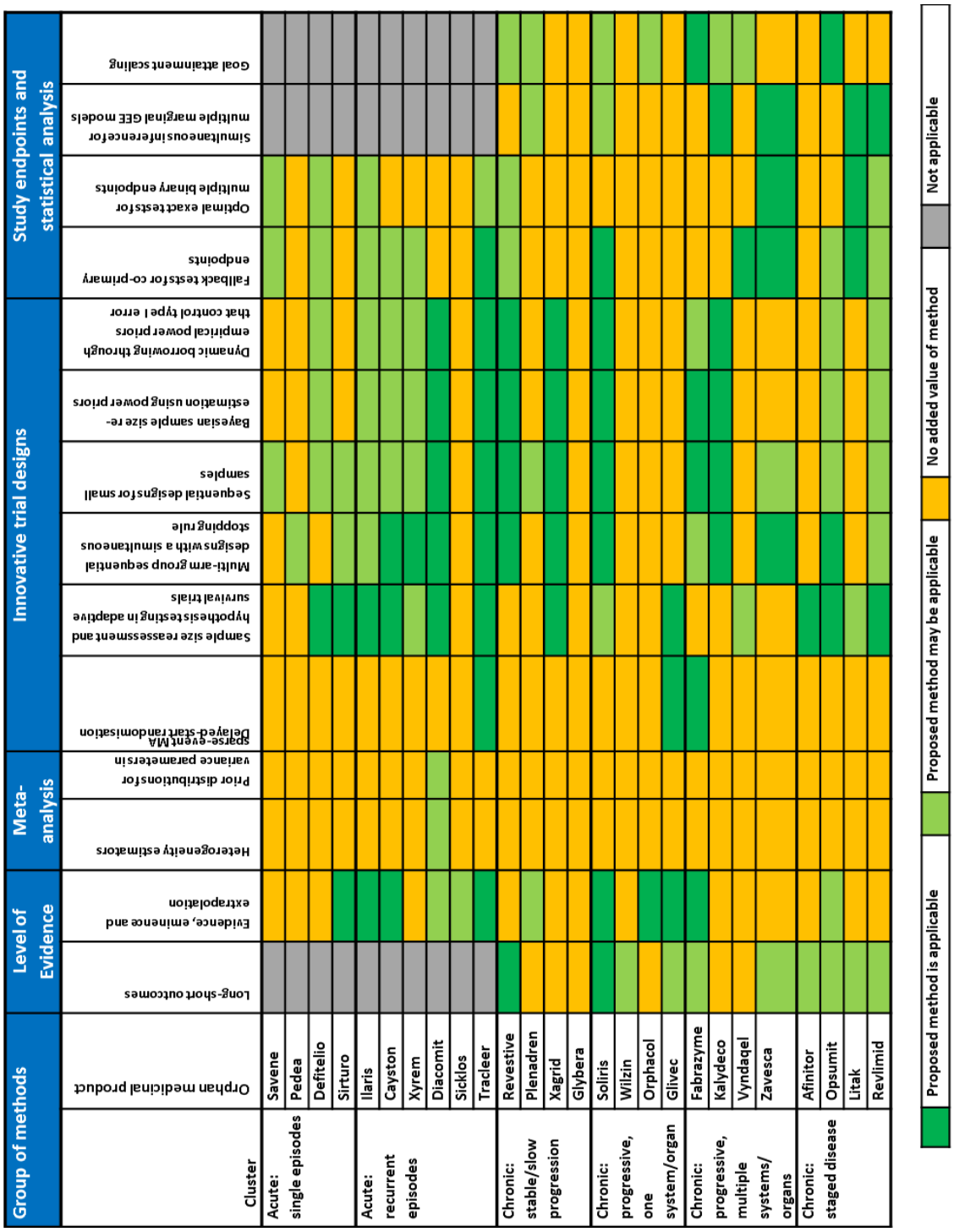


Figure 12. Heat map of applicability of novel ASTERIX methods.

1.1.3.7 General recommendations

ASTERIX focused on progress in clinical research for new treatments for *rare diseases*. The vision of ASTERIX is that such progress can be best made, by advancing *in coherence*: (1) statistical methodology for design and analysis, (2) incorporation of the patient perspective in design and outcomes and (3) uptake in practice and regulatory guidance after a careful assessment of information needs. For each of the main building blocks, critical advancement in methodology was achieved: for single trials, multiple trials and use of historical data, patient perspective and level of evidence. Uniquely, methods were evaluated against orphan drug development plans using European Public Assessment Reports and a comprehensive medical condition framework. Detailed recommendations for each of these areas are provided. In line with the coherent vision, we conclude with adding important, more general and overarching recommendations.

Statistical methodology for design and analysis

The progress in ASTERIX, as well as that in the companion projects IDeAl and InSPiRE, provides a hugely expanded statistical “tool box” to draw from, to improve clinical trials for small populations. ASTERIX specifically focused on rare diseases and the development of new treatments. From the totality of research, including the applications and the extensive evaluation against recent market authorisations, the following recommendation emerged. To optimise methodology for clinical development of new treatments for (very) rare diseases, the totality of the clinical research to be conducted needs to be considered. Simply optimising a single trial using some of the specific methods developed, may in fact be sub-optimal. Application of a carefully chosen array of methodologies, taking characteristics of the medical condition and treatment into account, will enable the best level of evidence for rare diseases.

Incorporation of the patient perspective in design and outcomes

ASTERIX has successfully involved patient representatives in a strongly methodological project. This has added value in the perception of all stakeholders involved. If this is the case for a project of this nature, it is evident that involvement of patient representatives needs to be considered always for projects related to clinical research, and needs to be done from the moment the project outline is conceived.

Uptake in practice and regulatory guidance

The (detailed) recommendations made by ASTERIX, as well as by IDeAl and InSPiRE, justify revision of the guideline for clinical trials in small populations. Importantly, application should take place at a much larger scale, although ASTERIX has clearly already applied the new methodology. It is crucial that the network can be sustained to support large scale application for industry sponsored as well as academic clinical research. This is likely to need continued public funding and public-private partnerships.

Finally, the joint projects ASTERIX, IDeAl and InSPiRE make a strong plea that sharing of patient level data be further advanced to make most efficient use of the limited information available on treatments for orphan diseases. The conduct and interpretation of multiple (post hoc) analyses in the context of regulatory decision-making is likely to need further expansion of bio-statistical expertise, to optimize procedural guidance and advanced analysis methods.

1.1.4 Potential impact, main dissemination activities and exploitation of results

1.1.4.1 Potential Impact

Ultimately, the positive impact of the "Advances in Small Trials dEsign for Regulatory Innovation and eXcellence" (ASTERIX) project should be to contribute to effective and safe new (drug) treatments reaching the many patients that suffer from a rare disease. Methodology for clinical trials has a key role to play, as it determines to a large extent both the efficiency of these trials as well as the reliability of the evidence derived from the results. Improved methodology may thus lead to quicker access to safe and effective treatments – as judged by all stakeholders - and protect patients from using ineffective but potentially harmful treatments. The ASTERIX consortium acknowledged that advancing clinical trial methodology is a multidisciplinary, multi-stakeholder effort. Progress in methodology of clinical research for new treatments for rare diseases was pursued by advancing in coherence statistical methodology, incorporation of the patient perspective and uptake in practice and regulatory guidance.

Science: Statistical methodology for design and analysis

The development of new innovative statistical methodology for single trials, series of trials and evidentiary standards is the backbone of the ASTERIX project. To assure that the new methodology could and would be implemented to the benefit of real clinical trials and drug development programs, the science first needs to reach clinical trial experts that implement trials.

Firstly, we have reached the clinical research community by publishing our scientific papers in both medical statistical, as well as more clinically oriented journals. Notably, the vast majority of papers is published open access – in line with ASTERIX explicit policy that we share our results between partners and with the broader community without any limits, in the spirit of open science. As the core program was driven by PhD projects, scientific output was very much concentrated near the end of the project. At this moment, there are 21 publications from the ASTERIX project in leading journals and another 20 either submitted or to be finished within the next three months (see references). Obviously, ASTERIX results have been published in medical statistical journals such as Statistics in Medicine (e.g. Urach et al.; 2016, Ristl et al., 2016) and Biometrical Journal (Koch, 2016). The recently accepted papers in Biometrics (Nikolakopoulos et al., 2017) and Statistics in Medicine (Pateras et al., 2017) exemplify that the research tackled fundamental approaches, from which further methodological developments will follow. ASTERIX also reached the clinical community through papers in Clinical Pharmacology and Therapeutics, PloS One (e.g. Magirr et al., 2016) and BMC Medical Research Methodology (Gaasterland et al., 2016) as well as journals specifically targeting rare diseases, like Orphanet Journal of Rare Diseases (e.g. Hilgers, Roes and Stallard, 2016; Derhaschnig et al. 2016; Lasch 2017). Applications to diseases in need of new treatments are also finding their way to the clinical scientific journals (e.g., van Eijk, 2017; amyotrophic lateral sclerosis and in general neurodegenerative disorders).

Secondly, the novel methodology and scientific evaluations have shown the (proven) potential to improve efficiency and reliability of clinical trials or clinical drug development programs. The extensive evaluation of methods against European Public Assessment Reports has made this potential clear, on top of the many applied examples that were driving the methodological work. Part of the novel methods will lead to what could be seen as *incremental improvements in clinical research*. Improvements will entail deriving (much) more information from limited sample sizes, reducing the total time needed for a clinical trial or

preventing waste by stopping (part) of a trial early in case no benefit is likely to come from it. Thus, methods both increase efficiency as well as the added value (information) from small population trials. Just a few examples can illustrate this. The methods for optimising sequential designs conditional on a maximum achievable sample size and adaptive designs with survival endpoints allow early decision making, tailored to maximum use of information as well as availability of only small samples. The fact that the proposed sequential optimisation is already used as trial design in amyotrophic lateral sclerosis exemplifies this potential. The fallback procedure and exact tests for leveraging information from multiple endpoints both increase efficiency, while being tailored to small samples.

Importantly, ASTERIX has also made progress that can be considered more *breakthrough* by itself, or facilitating truly different approaches to clinical trials and development for rare diseases. The advancement of Goal Attainment Scaling as patient centred outcome instrument, will allow a more heterogeneous patient population being entered into clinical trials. This is a two-sided sword: it will not only help recruiting more patients, it will facilitate more rare disease patients potentially profiting from new treatments, expressed in goals that are important to them. As part of ASTERIX collaboration with Khondrion, Goal Attainment Scaling was already added to a clinical trial investigating a new drug for mitochondrial diseases, whose results are due the coming months. The advancement in adaptive multi-arm trials and the example of a basket trial can support drug development strategies that are especially relevant in rare diseases. In view of the (very) small target population, there is a strong imperative to avoid separate clinical trials for different treatments in the same population. Proper methodology such as presented, facilitates different treatments being investigated in the same trial, amongst other uses. Thus, methodology will help to change the development paradigm, with different pharmaceutical companies join forces in ways that seemed far-fetched at the start of ASTERIX. The investigation into meta-analysis in small series of small trials as well the prior data conflict calibrated power priors have shed fundamental light on how to perform and assess analyses across trials. This can impact design of trials, and make series of trials more efficient. The deeper insights obtained also improve (regulatory) assessment for any integrated data analyses in the small series of small trials setting.

Thirdly, the scientific development provided *more fundamental insights* and may drive new avenues of methodology and *stimulate further development*. This extends impact beyond the scope of ASTERIX and possibly beyond the scope of rare diseases. Many of these more fundamental results pertain to the *level of evidence* needed for assessment from a clinical, regulatory or reimbursement perspective. The EU legislation determines that market access to new drugs requires the same level of evidence regardless of whether they are intended for (very) rare or highly prevalent diseases (Regulation (EC) No 141/2000). Nevertheless, actual decision making does seem to do so, but systematic guidance is lacking. Better understanding when and how more relaxed levels of evidence could be allowed in the regulatory decision making has major impact. Although a final understanding and consensus of all aspects of the decision problem and its solution may not have been reached yet, ASTERIX research has been instrumental in the ongoing discussions on this critical topic. The combined work on randomisation, stratification and meta-analysis is a case in point. The impact of between study or, equivalently, between strata heterogeneity is such, that clear cut solutions for sufficiently powered analyses in the presence of heterogeneity do not exist in case of a small number of strata (of small size). This not only impacts small population clinical research, but all trials with a small number of strata. Although ASTERIX was able to provide Bayesian approaches that are at least somewhat robust, further research is needed.

ASTERIX research further supported with real data that *randomisation* as early as possible in the clinical development program is likely to be preferred, both from an overall efficiency point of view (smaller number of patients), as well as for consistency of the evidence generated (see Figure 10). There may, however, be situations where

randomisation may not be possible in an early stage. The concept of threshold-crossing as initiated by the EMA and developed by IDeAl in collaboration with ASTERIX, provides a viable approach for this setting, where of course measures can and should be taken to reduce the potential (large) selection bias in the single arm trials. Both research lines have added valuable insights into determining the optimal strategy, but this does require further discussion. Especially in very rare diseases, single arm trials and disease registries are more common as evidence base, and ASTERIX has provided key research approaches to further determine the proper use and assessment of this type of evidence.

The approach to use prior data conflict calibrated power priors as a means to incorporate prior data into the design of a trial can be extended in many directions. It thus provides a more generic approach that can increase efficiency and consistency of evidence in many settings that can profit from including historical data.

Finally, the joint work with IDeAl on Evidence, Eminence and Extrapolation provides a framework based on prior beliefs to reduce the size of the drug development programme in a small population. The prior beliefs needed for extrapolation to a small target population – such as paediatric or orphan indications - are based on prior information from clinical trials in a related larger source population. This approach was enthusiastically received by EMA, and fits within the draft EMA guidance on extrapolation. It is thus likely to have substantial impact, whereas it is also clear that more detailed development of the framework is needed for application in specific situations.

Patient perspective

First and foremost, the fact that ASTERIX included an active *Patient Think Tank* (Table 1), ensured a process of constant feedback from the patient perspective. In particular, this Patient Think Tank has actively contributed to the development of the POWER model of patient involvement in trial design. It has been active during the entire project and had high impact on the constant awareness of ASTERIX researchers (mainly statisticians) that new methodology had to be relevant for patients in the end. As such they have trained all ASTERIX researchers, and in particular had positive impact on the training of the young statisticians.

Rare disease	Patient representative
Cancers - Inspire2Live, Netherlands	Ms. Veronica van Nederveen
Duchenne Muscular Dystrophy: Duchenne Parents Project, Netherlands.	Ms. Elizabeth Vroom
Mucopolysaccharidoses, MPS Society, UK	Ms. Christine Lavery
Acute Myeloid Leukaemia, Ireland	Ms. Nóirín O'Neill
Find a Cure & Alkaptonuria, AKU Society, UK	Mr. Oliver Timmis
European Hemophilia Consortium, Poland	Mr. Radoslaw Kaczmarek
Primary Sclerosing Cholangitis, PSC International, Netherlands	Ms. Marleen Katee
Fragile X Support Group, Germany	Mr. Jorg Richstein
Alström Syndrome UK, & Eurordis	Ms. Kerry Gleeson-Beevers

Table 1. ASTERIX Patient Think Thank members

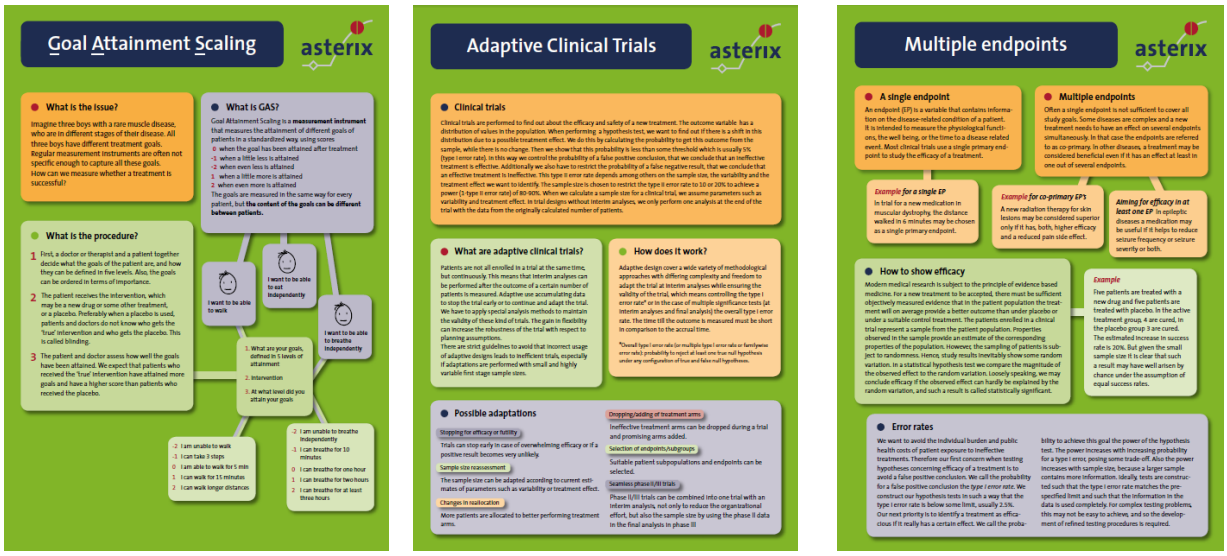
The fact that ASTERIX included the patient representatives successfully has paved the way for future methodology projects. To support this, involvement of the patient representatives has also been clearly visible – through co-authoring papers, and presenting (jointly) at international conferences (Picture 2).

Picture 2. Joint session at SCT of researchers and patient representative.

Society of Clinical Trials International Clinical Trials Methodology Conference, May 2017. Kit Roes, Hanneke van der Lee, Charlotte Gaasterland and Susane Urach as ASTERIX researchers and Patient Representative Alex Johnson. She gave a very inspiring talk on her experiences as a mother of a 9-year old son with Duchenne.

Among the patient centred research results, Goal Attainment Scaling methodology was very well received as indicated above. To increase its potential impact, the consortium considers submitting the methodology for qualification advice to EMA to be approved as acceptable methodology.

The patient think tank and ASTERIX researchers have created *Patient Information Leaflets* (Picture 3) to translate results on relevant topics for patients and patient organisations into laymen terms. Together with the increased demands by regulations to provide layman terms summaries of clinical study results, the patient information leaflets will reach out to patient communities to have up to date insights and better understanding of innovations in clinical developments.



Picture 3. Examples of Patient Information Leaflets.

This is expected to support the dissemination and application of the new proposed methodology to other patient organisations. These Patient Information Leaflets are available on the Asterix website. http://www.asterix-fp7.eu/patient-groups/leaflets/.

Regulatory impact and uptake in practice

Novel methodology for clinical trials will only have impact for patients when implemented in clinical trials in practice and if it is acceptable for regulators. Through the composition of the ASTERIX consortium partners and external Advisory Board members, the ASTERIX project ensured that potential implementation in practice and acceptability by regulatory agencies was optimised.

Partner		Lead
Universitair Medisch Centrum Utrecht	University Medical Center Utrecht *EMA Biostatistics Working Party:*	Prof Dr Kit Roes *Observer*
MHH Medizinische Hochschule Hannover	Medizinische Hochschule Hannover *EMA Biostatistics Working Party:*	Prof Dr Armin Koch *Member*
MEDICAL UNIVERSITY OF VIENNA	Medizinische Universität Wien *EMA Biostatistics Working Party:*	Prof Dr Martin Posch *Observer*
amc	Academic Medical Center Amsterdam	Dr Hanneke van der Lee
UAB Universitat Autònoma de Barcelona	Universitat Autònoma de Barcelona *EMA Biostatistics Working Party:*	Dr Ferran Torres *Member*
PSI	Statisticians in Pharmaceutical Industry	Dr Egbert Biesheuvel
VSOP	VSOP (Vereniging Samenwerkende Ouder- en Patiëntenorganisaties Betrokken Bij Erfelijkheids-vraagstukken)	Dr. Cor Oosterwijk

Table 2. ASTERIX Consortium work package leaders with their connections to EMA.

The deep involvement of consortium members in regulatory work (Table 2), both in their respective countries as well as at EMA level, ensures short- and long-term impact on regulatory uptake of new methods and insights. It is already incorporated in their contributions to Scientific Advice and assessments, many of which are actually on orphan designated drugs. It also clearly facilitated that revision of the guideline for clinical trials in small population was put on the agenda for 2018. Most likely, consortium members in the Biostatistics Working Party will be asked to take the lead in this. In addition, these members are also involved in numerous more disease specific challenges and guidelines, which provides additional opportunity to tailor methodology from ASTERIX, as well as the sister projects IDeAl and InSPiRe, to concrete drug development challenges.

Moreover, all scientific consortium partners are at academic medical centres. Thus, many of the methodological approaches were inspired by real life challenges in clinical research for rare diseases. The medical environment also enabled them to ‘test drive’ the

methods they were developing. This is the case for muscle and neurodegenerative diseases such as amyotrophic lateral sclerosis, mitochondrial diseases (goal attainment scaling), cystic fibrosis (trial design), and transplantation research and oncology. Novel trial design for rare diseases is, e.g., now part of an approved H2020 clinical research project to evaluate (within one trial) drugs from different pharmaceutical companies to treat patients with very rare to ultra-rare genetic variants of cystic fibrosis. Additionally, Profs Koch, Posch and Roes are involved as key experts in **connect4children**, the proposed collaborative network for European Clinical Trials for Children. This network will generate a sustainable infrastructure that optimises the delivery of clinical trials in children. The role of the ASTERIX experts is to provide high quality input to study design and to promote the use of innovative methodologies. Of course, to leverage the excellent collaboration so far, experts from IDeAl and InSPiRE will be connected as well.

Next generation, next steps – beyond ASTERIX

Long term impact of ASTERIX will be materialized by the fact that a new generation of young statisticians was trained. With 4 PhDs completed, and an ultimate total number of about 10 expected, a sizeable group of new experts was created. They can continue to add value in their respective institutes, such as Susanne Urach (MUW). She was awarded with the L'Oreal Austria Fellowship which will enable Susanne to continue her academic career as a postdoc a the Medical University of Vienna. Others will follow the example of Stavros Nikolakopoulos (UMCU) who obtained his PhD late 2016, and moved to Athens as a next career step – taking with him the experience gained.

The principal investigators of ASTERIX, IDeAl and InSPiRE have committed themselves to sustaining and growing the joint network and make sure methods are implemented to the benefit of patients suffering from a rare disease. They made this commitment at the "Seventh Framework Program (FP7) small-population research methods projects and regulatory application workshop", March 29&30, 2017, which was jointly organised by the European Medicines Agency (EMA) and the three EU projects. Concretely, they are shaping a European training network to make that happen.

1.1.4.2 Main dissemination activities

At the start of the project a clear dissemination plan was set, that included diverse routes of communication and stakeholder engagement. The external dissemination was aimed to increase the visibility and awareness of the ASTERIX project and to disseminate all results to the scientific community, patient communities and regulators. All project partners engaged actively in dissemination activities.

The *internal dissemination* approach, involving all researchers, advisory boards and Patient Think Tank has led to fruitful collaboration across the network, including scientists exchanges. It is also reflected in joint papers, and a truly ASTERIX-wide joint evaluation of methods against regulatory experience. It has created a firm base to sustain the network beyond the project level.

Of most interest to patients and society, is the level and success of *external dissemination*. From a project like this, benefits for patients can only be achieved if progress and results reach application by stakeholders such as clinical researchers, industry and regulatory authorities. Key external dissemination channels consisted of the ASTERIX website, scientific publications, presentations at conferences, promotion materials and interactions with regulators and the Patient Think Tank. In addition, we were keen on any opportunity to contribute to advancement of clinical research for rare diseases.

Highlights of the direct results of our external dissemination activities include:

- Scientific output in over 40 papers, of which 21 published in major journals and the remaining expected within the next 3 months. More than half open access, reflecting the open science strategy.
- Over 100 presentations and invited sessions in conferences and workshops, with a substantial number aimed at a broader clinical and regulatory audience and some by or jointly with patient representatives.
- A firm partnership with patient organisations.
- A firm partnership between the three projects ASTERIX, IDeAl and InSPiRE.
- Tangible results in the form of Patient Information Leaflets to share results.
- Providing expert input for clinical trial design in several major H2020 projects in which clinical research is conducted for rare diseases or children.
- Membership of Steering Committee of the IRDiRC task force on small population clinical trials, co-authoring the IRDiRC report.
- Direct input in updating regulatory guidance, as well as regulatory advice and assessment.

Scientific publications

In total 21 publications in leading journals are (co)authored by ASTERIX researchers to reach the science community, both statistical and clinical, as well as the rare disease community. It is worth noticing that the vast majority of these publications have open access. At the moment of writing this report, a similar number of publications is submitted or under review at scientific journals. The relative concentration of papers completed near the end of the project is a logical consequence of the focus on PhD projects. The scope and impact of these papers is described in the previous chapters of this report.

Presentations and sessions at conferences

ASTERIX researchers have been presenting their results at many conferences, workshops and international events, with in total more than 100 presentations and 50 posters. As could be expected the majority of the scientific presentations and posters were presented at medical statistical events. In addition to the scientific value of presenting results, this also increased awareness of the crucial importance to develop methodology that would specifically help rare disease clinical research. The 2015 conference of the International Society of Clinical Biostatistics presents an illustrative example. The principal investigator Prof. Kit Roes of ASTERIX was the chair of the Scientific Program Committee. At the conference, rare disease clinical research was a conference wide theme, with invited sessions and a key note lecture by Rob Hemmings (Medicines and Health Regular Authority from the UK and chair of the Scientific Advice Working Party of the EMA). It further resulted in a special topic issue of Biometrical Journal (Roes, 2017) and an agreement to carry rare disease clinical research as one of the key themes for the next couple of conferences as well. The principal investigator of InSPiRE, Prof. Nigel Stallard was subsequently the chair of the 2016 conference, and also the 2017 ISCB meeting in Vigo, Portugal, hosted a substantial number of small population methodology presentations and posters. The 2015 conference was in addition used to further strengthen the collaboration between ASTERIX, IDeAl and InSPiRE (Picture), facilitated by the fact that researchers from all three projects had a presentation or poster at the conference.

At the Joint Statistical Meeting 2016 in Chicago, Susanne Urach (Medical University of Vienna) received the Student Paper Award for her paper "Multi-arm Group sequential Designs with a Simultaneous Stopping Rule" (Urach, 2016).

The latest major statistical conference ASTERIX researchers contributed to, was the Joint meeting on Biometrics and Biopharmaceutical Statistics (CEN – ISBS) in Vienna, August 2017. At this global conference, experts from statistics and epidemiology across the world gathered. In addition to presentations of the different individual research topics from ASTERIX, a separate invited session was dedicated to share the results of the ASTERIX, IDeAl and InSPiRE. The session was organised by Prof. Martin Posch, chair of the Scientific Program Committee and involved also the DIA Small Populations Working Group. Susanne Urach and Robin Ristl received the Arthur Linder Prize of the Austrian Swiss Region of the International Biometric Society (IBS) at this meeting.

Research and results were also shared and discussed at more general clinical and pharmaceutical conferences. These included the annual conference on pharmaceutical development (FIGON, 2016) in the Netherlands, and the European Conference on Rare Diseases & Orphan Products (2016) in Edinburgh, among several European and national conferences (see website). At the joint Society of Clinical Trials and International Clinical Trials Methodology Conference, May 2017. Kit Roes and Hanneke van der Lee chaired an invited session on patient centred outcomes, that stirred fruitful discussion. The session also included Patient Representative Alex Johnson, who gave a very inspiring talk on her experiences as a mother of a 9-year old son with Duchenne.

Finally, of course the ASTERIX Final Symposium in September 2017 (Zaandam, the Netherlands) is worth highlighting. In line with the overall approach, it followed the multi-stakeholder format by involving and integrating the views of patients, regulators and clinical researchers in sessions on the role of randomization, meta analyses, alternative endpoints, level of evidence, and on ethical framework. (http://www.asterix-fp7.eu/results/end-symposium/)

Picture 4. From left to right: Edwin Spaans (industry), Elizabeth Vroom (Patient Representative) and Leonard van den Berg (clinical investigator).

Promotion materials and interactions with the Patient Think Tank

The most important promotion materials produced are the Patient Information Leaflets described above (see Picture 3). In addition, the open science approach ensures that presentations, posters and any further products (such as R packages) are openly shared. The website will continue to be maintained and updated by the coordinator's institute, to ensure continuity after the project as such is finished.

Networking and other opportunities

Another important way to disseminate results with impact is networking with many different stakeholders, including the "joint projects" IDeAl and InSPiRE, regulators, the rare disease community, industry and many more.

The three projects organised the joint Small Population Symposium in Vienna in 2014 – as a kick-off. Following from that, the three principal investigators also published a shared view on 'directions for new developments on statistical design and analysis of small population group trials' (Hilgers, Roes and Stallard, 2016).

On behalf of ASTERIX, Prof Martin Posch was invited speaker and discussant on adaptive designs at various FDA and EMA workshops in 2014. Researchers from ASTERIX and IDeAl were invited at the Japanese PMDA in March 2015 to discuss trends in adaptive designs and statistical approaches to small population group trials in the European Union. (http://www.pmda.go.jp/files/000204874.pdf).

Left: Discussion at the Q & A session. Right: from the left Dr. Bauer, Dr. Posch and Dr. König,

Similarly, dr Hanneke van der Lee participated in a workshop on outcome measurement in rare diseases, at Duke-Margolis Center for Health Policy, involving academia, regulators (FDA) and industry.

An important meeting, organized by IRDiRC at the EMA in London, hosted invited experts to discuss actions to reach agreement between the different stakeholders on appropriate small population studies (March 3rd, 2016). On behalf of ASTERIX, dr Ferran Torres and Profs Martin Posch and Kit Roes (who is also member of the IRDiRC Steering Committee on the topic) participated in the discussions. The extensive discussion resulted in recommendations, that were laid down in an IRDiRC report and are submitted for publication (Jonker et al. 2016).

A year later, the collaboration among the three projects resulted in the "Seventh Framework Programme (FP7) small-population research methods projects and regulatory application workshop", which was jointly organised by the EMA. At this enticing meeting it was clear that the EMA is supportive of innovative methods in this field, and would urge to make application concrete. It was considered that part of the results are ready for broad(er) implementation, which can be stimulated by regulatory instruments such as the EMA Guidance on Small Populations, and the EMA Qualification Procedure. Both are pursued by ASTERIX, as described above. Due to ASTERIX, patient representatives participated as well and expressed their appreciation for the opportunity to truly engage actively in advancing

research methodology for clinical trials. The live stream of the workshop (held March 29-20, 2017) was followed globally. The full content is available on line (see www.ema.europa.eu/ema/ under News and Events).

Most recently, Prof Kit Roes presented and discussed at the DIA/EFGCP/EMA Conference on How to Optimise Children's Access to Innovative Medicines (London, Oct 16&17, 2017). It provided the opportunity to connect the methodological research of all three FP7 projects to the initiatives to accelerate and improve clinical drug research for children.

Finally, interaction with industry to increase awareness and enhance implementation was ensured by the two representatives in the Advisory Board as well as PSI/EFSPI being a partner in the ASTERIX consortium. As a result, EFSPI has set up a new Special Interest Group for Small Populations early 2016. This Small Population SIG is a forum for statisticians, mainly from pharmaceutical industry, to discuss new methodology and exchange information and best practise situations during monthly webinars. *(http://www.efspi.org/EFSPI/Special_Interest_Groups/Small_Populations_SIG.aspx).*

1.1.4.3 Exploitation of results

Exploitation of methodological results is best realized by sharing in an open way. Already at the conception of ASTERIX and the start of the project, all partners were convinced that results should be freely available, for everybody to use. It is our belief that sharing is the new way of multiplying, which is aligned with the multi-disciplinary structure and multi-stakeholder networking approach of the whole project. This will be supported by 'open science', sharing results in an open way to allow access for all and thereby broad and easy implementation. Most publications of ASTERIX have open access and also future publications will have open access where possible. The website will contain all presentations and papers, and will continue to be maintained and updated.

The second form of exploitation is to implement methodology in actual clinical research and regulatory work. The ASTERIX consortium is uniquely positioned to do so, combining strong regulatory involvement with direct engagement of researchers in clinical trials at their respective academic medical centers. The fact that the European Federation of Statisticians in the Pharmaceutical Industry is part of consortium and has formed a Special Interest Group, ensures a strong connection with application in industry. Current and expected results are described under 1.1.4.1 Potential impact. The major initiatives (H2020 clinical projects, connect4children, European training network) have tremendous potential.

The continued collaboration between the three principle investigators of ASTERIX, IDeAl and InSPiRE underlying these initiatives provides a unique European network to build on. All three leading coordinators are committed to continue working together to enhance applying innovative methodology in actual trials as well as guidance documents, facilitating training needs and continue efforts to develop new innovative methodology.

References

Austin P., H.J.s. Dawkins (2017).Medical research: Next decade's goals for rare diseases. *Commentary. Nature,* 548, 158 (10 August 2017), doi:10.1038/548158c.

Brakenhoff T.B., K.C.B. Roes, S. Nikolakopoulos (2017). Bayesian sample size re-estimation using power priors. *Submitted Stat. Meth. in Med Res.*.

CHMP/EWP/233/99. Points to Consider on application with 1. Meta-analysis; 2: One pivotal study. London: EMA, 2001.

CHMP/EWP/83561/2005. Guideline on clinical trials in small populations. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf, London: EMEA, 2006.

Derhaschnig, Ulla, Jim Gilbert, Ulrich Jäger, Georg Böhmig, Georg Stingl, Bernd Jilma (2016). Combined integrated protocol/basket trial design for a first-in-human trial. *Orphanet Journal of Rare Diseases*, DOI 10.1186/s13023-016-0494-z).

Eichler, H. G., Bloechl-Daum, B., Bauer, P., Bretz, F., Brown, J., Hampson, L. V., Posch, M., F Koenig (2016). "Threshold-crossing": A Useful Way to Establish the Counterfactual in Clinical Trials? *Clinical Pharmacology & Therapeutics*, 100(6), 699-712.

van Eijk R.P.A, Stavros Nikolakopoulos, Toby Ferguson, Dawei Liu, Marinus J.C. Eijkemans, Leonard H. van den Berg (2017). Increasing the efficiency of clinical trials in neurodegenerative disorders using group sequential trial designs. *Submitted to JAMA Neurology.*

Emanuel EJ, Wendler D, Grady C. (2008). An ethical framework for biomedical research. In: Emanuel et al ed. The Oxford Textbook of Clinical Research Ethics. OUP 2008, pg 123-135.

Gaasterland C.M.W., M.C. Jansen-van der Weide, E. Vroom, K. Leeson-Beevers, K. C. B. Roes, J.H. van der Lee (2016a). The POWER model: recommendations for involving patient representatives in choosing relevant outcome measures during clinical trial design. *To be submitted.*

Gaasterland CMW, Jansen-van der Weide MC, Weinreich SS, van der Lee JH. (2016b). A systematic review to investigate the measurement properties of goal attainment scaling, towards use in drug trials. *BMC Med Res Methodol*. 2016;16:99.

Gaasterland, C.M.W., Jansen-van der Weide, M.C., Roes, K.C.B., van der Lee, J.H. (2017). Goal Attainment Scaling as an outcome measure in rare disease trials: A blueprint for validation. *To be submitted.*

Galwey N.W. (2016). Supplementation of a clinical trial by historical control data: is the prospect of dynamic borrowing an illusion? *Statistics in Medicine*. Published online 7 December 2016 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/sim.7180.

Gonnermann, A., Framke, T., Großhennig, A., & Koch, A. (2015). No solution yet for combining two independent studies in the presence of heterogeneity. *Statistics in Medicine*, http://doi.org/10.1002/sim.6473.

van der Graaf R., Kit CB Roes, Sofieke Dekker, Egbert Biesheuvel, Martine Janssen-van der Weide, Johannes JM van Delden (2017). Ethical benchmarks for rare diseases research. *To be submitted.*

Hasford J., A. Koch (2017). Ethical aspects of clinical trials in rare diseases. *Bundesgesundheitsbl* 2017 · 60:556–562 (in German).

Hilgers Ralph-Dieter, Kit Roes, Nigel Stallard (2016). Directions for new developments on statistical design and analysis of small population group trials. *Orphanet Journal of Rare Diseases* (2016) 11:78. DOI 10.1186/s13023-016-0464-5.

Hlavin, G., Koenig, F., Male, C., Posch, M., & Bauer, P. (2016). Evidence, Eminence and Extrapolation. *Statistics in medicine*, 35(13), 2117-2132.

Hofer, M.P. Hanna Hedman, Maria Mavris, Franz Koenig, Thorsten Vetter, Martin Posch, Spiros Vamvakas, Jan Regnstrom and Stiina Aarum (2017). Marketing authorisation of orphan medicines in Europe from 2000 to 2013, *Drug Discov Today*, https://doi.org/10.1016/j.drudis.2017.10.012.

Jansen-van der Weide M.C., Charlotte C.M. Gaasterland, Kit C.B. Roes, Caridad Pontes, Roser Vives, Arantxa Sancho, Stavros Nikolakopoulos, E. Vermeulen, Johanna H. van der Lee1 (2017). Rare disease registries: potential applications towards impact on development of new drug treatments. *To be submitted.*

Joergens Silke, Gernot Wassmer, Franz Koenig, Martin Posch (2017). Confirmatory Adaptive Designs with a Survival Endpoint Using a Short-Term Endpoint for Treatment Arm or Population Selection. *In preparation*.

Jonker AH, Mills A, Lau LPL, Ando Y, Baroldi P, Bretz F, Burman CF, Collignon O, Hamdani M, Hemmings RJ, Hilgers RD, Irony I, Karlsson M, Kirschner J, Krischer JP, Larsson K, Leeson-Beevers K, Molenberghs G, O'Connor D, Posch M, Roes KC, Schaefer F, Scott J, Senn SJ, Stallard N, Thompson A, Torres F, Zohar S, Aymé S, Day S Eds. (2016). "Small Population Clinical Trials: Challenges in the Field of Rare Diseases." IRDiRC Workshop Report and Recommendations.

Koenig, F., Slattery, J., Groves, T., Lang, T., Benjamini, Y., Day, S., and Posch, M. (2015). Sharing clinical trial data on patient level: opportunities and challenges. *Biometrical Journal*, 57(1), 8-26.

Koch, A., & Ziert, Y. (2016). Does it help that efficacy has been proven once we start discussing (added) benefit? *Biometrical Journal*, 58(1), 89–103. http://doi.org/10.1002/bimj.201400017

Kunz, CU, Friede T, Parsons N, Todd S and Stallard, N. (2014). Data-driven treatment selection for seamless phase II/III trials incorporating early-outcome data, *Pharmaceutical Statistics*, 2014; 13 238–246.

Lasch F., Weber K., Chao M. M., Koch A. (2017) A plea to provide best evidence in trials under sample-size restrictions: the example of pioglitazone to resolve leukoplakia and erythroplakia in Fanconi anemia patients. *Orphanet Journal of Rare Diseases,* 2017 12:102, DOI: 10.1186/s13023-017-0655-8.

Magirr, D., Jaki, T. F., Koenig, F., Posch, M. (2015). Sample size re-assessment and hypothesis testing in adaptive survival trials. *PLoS ONE*, 11(2): e0146465.

Musuamba, F. T., Manolis, E., Holford, N., Cheung, S. Y. A., Friberg, L.E., Ogungbenro, K., Posch, M. & Corriol-Rohou, S. (2017). Advanced Methods for Dose and Regimen Finding During Drug Development: Summary of the EMA/EFPIA Workshop on Dose Finding (London 4–5 December 2014). *CPT: pharmacometrics & systems pharmacology*, 6(7), 418-429.

Nikolakopoulos S., van der Wal W.M., Roes K.C.B. (2013). An analytical approach to assess the predictive value of biomarkers in Phase II decision making. *Journal of Biopharmaceutical Statistics*, 23, p. 1106-1123.

Nikolakopoulos S., K.C.B. Roes, and I. van der Tweel (2016). Sequential designs with small samples: Evaluation and recommendations for normal responses. *Stat. Meth. in Med. Res.* Article first published online: June 24, 2016. *https://doi.org/10.1177/0962280216653778.*

Nikolakopoulos S., I. van der Tweel and K.C.B. Roes (2017). Dynamic borrowing through empirical power priors that control type I error. *Accepted for publication, Biometrics*.

Oude Rengerink K., Mitroiu M. , Pontes C., Sancho A, Vives R. Pesiou S., Torres F., Rosenkranz G., Posch M., Urach S., Ristl R., Koch A., Spineli L., Nikolakopoulos S., Pateras K., van der Lee J.H., van der Weide J.M., Gaasterland C.M.W., Roes K.C.B. (2017). Applicability and added value of novel methods to improve drug development in rare and ultra-rare diseases. *To be submitted.*

Pateras K., Nikolakopoulos S. and Roes K.C.B (2017). Data generating models of dichotomous outcomes: Heterogeneity in simulation studies for a random-effects meta-analysis (*Accepted for publication, Stat in Medicine*).

Pateras K., Nikolakopoulos S., Mavridis D. and Roes KCB (2017). Interval estimation of the overall treatment effect in a meta-analysis of a few small studies with zero events (under review, *Contemporary Clinical Trials Communications*).

Pateras K., Spineli L. and K.C.B. Roes (2017). Bayesian evidence synthesis for combining results of series of a few small trials. *To be submitted*.

Pateras K., Nikolakopoulos S. and Roes K.C.B. (2017). Prior distributions for variance parameters in a sparse-event meta-analysis of a few small trials *To be submitted*.

Pontes C., R. Vives, M-G Valent, M. Fontanet, J. Rios, R. Morros, A. Sancho, J. Martinalbo, M. Posch, A. Koch, K.C.B. Roes, J. Torrent-Farnell, F. Torres (2017). Proposal of rare medical condition clustering to guide clinical development of orphan drugs. *To be submitted.*

Ristl, R., Frommlet, F., Koch, A., Posch, M. (2016). Fallback tests for co-primary endpoints. *Statistics in Medicine*, 35(16), 2669-2686.

Ristl R, Urach S, Rosenkranz G, Posch M (2017). Methods for the analysis of multiple endpoints in small populations: A review. *Journal manuscript in revision*.

Ristl R, Xi D, Glimm E, Posch M, (2016). Optimal exact tests for multiple binary endpoints ARXIV https://arxiv.org/abs/1612.07561. *Journal manuscript in revision at Computational Statistics and Data Analysis.*

Ristl R, Ritz C, Posch M, Hothorn LA, (2017). Simultaneous inference for multiple generalized linear models with dependent observations. *In preparation*.

Roes KC. (2016). A framework: make it useful to guide and improve practice of clinical trial design in smaller populations. BMC Med. 2016 Nov 25;14(1):195.

Roes K. (2017). Editorial on special topic ISCB 2015: "Clinical research design, rare diseases, and personalized medicine". Biom J. 2017 Jul; 59(4):607-608. doi:10.1002/bimj.201700065.

Senn S. (2004). Added values. Controversies concerning randomization and additivity in clinical trials. *Statist. in Med*. 23:3729–3753.

Spineli LM, Jenz E, Großhennig A, Koch A. (2017). Critical appraisal of arguments for the delayed-start design proposed as alternative to the parallel-group randomized clinical trial design in the field of rare disease. *Orphanet J Rare Dis*. 2017 Aug 17;12(1):140.

Stallard N and Todd S (2010). Seamless phase II/III designs, *Statistical Methods in Medical Research*, 2010 ; 20(6) 623–634.

Urach, S., Posch M. (2016) Multi-arm Group Sequential Designs with a Simultaneous Stopping Rule. *Statistics in Medicine*, DOI:10.1002/sim.7077.

Urach S., Gaasterland C.M.W., Posch M., Jilma B., Roes K., Rosenkranz G., Van der Lee J.H., Ristl R. (2017). Statistical analysis of Goal Attainment Scaling endpoints in randomised trials. *Submitted Statistical Methods in Medical Research*.

Urach, S., Posch M. (2017), Adaptive Parametric Multiple Testing Procedures. *In preparation.*

Zafiropoulos, Nikolaos, Franz Koenig, Francesco Pignatti, Andreas Kouroumalis, Lorenzo Guizzaro, Kit Roes, Ferran Torres, Martin Posch (2017). Uncertainties and coping strategies in the regulatory review of orphan medicinal products. I*n preparation.*

