Universitat Autònoma de Barcelona

New clinical trial designs applied to the study of orphan and rare diseases: feasibility of methodological guidance to clinical development of new treatments from a regulatory perspective

ARANZAZU SANCHO LÓPEZ

Universitat Autònoma de Barcelona

DOCTORAL THESIS

New clinical trial designs applied to the study of orphan and rare diseases: feasibility of methodological guidance to clinical development of new treatments from a regulatory perspective

-ARANZAZU SANCHO LÓPEZ-

Barcelona, 2018

Universitat Autònoma de Barcelona

DOCTORAL THESIS

New clinical trial designs applied to the study of orphan and rare diseases: feasibility of methodological guidance to clinical development of new treatments from a regulatory perspective

-ARANZAZU SANCHO LÓPEZ-

Department of Pharmacology, Therapeutics and Toxicology
PhD in Pharmacology

THESIS DIRECTORS:

Dr Caridad Pontes García
Unitat Docent Parc Taulí,
Facultat de Medicina, UAB

Dr Ferran Torres Benítez
Unitat Bioestadística, Hospital Clínic
Facultat de Medicina, UAB

TUTOR:
Dr Marta Valle Cano
Institut de Recerca HSCSP
Hospital de la Santa Creu i Sant Pau

Barcelona, 2018

Universitat Autònoma de Barcelona

Department of Pharmacology, Therapeutics and Toxicology

Caridad Pontes García, Associate professor, Department de Farmacologia, de Terapèutica i de Toxicologia, Unitat Docent Parc Taulí, Facultat de Medicina - Universitat Autònoma de Barcelona,

Ferran Torres Benítez, Unitat Bioestadística. Departament de Pediatria, Ginecologia, Obstetrícia i Medicina Preventiva, Facultat de Medicina- Universitat Autònoma de Barcelona,

Marta Valle Cano, Associate professor, Department de Farmacología, de Terapèutica ide Toxicologia. Universitat Autònoma de Barcelona,

CERTIFY: that the present doctoral thesis, titled “New clinical trial designs applied to the study of orphan and rare diseases: feasibility of methodological guidance to clinical development of new treatments from a regulatory perspective”, has been performed by D. Aránzazu Sancho Lopez under our supervision.

Barcelona, 9th August 2018

Signatures:

Caridad Pontes García

Ferran Torres Benítez

Marta Valle Cano

Gracias a la vida que me ha dado tanto
…

Violeta Parra

Con todo mi agradecimiento a Caridad Pontes y Ferran Torres, mis directores y, sobre todo, mis amigos. Sin su iniciativa, entusiasmo, motivación y su dedicación este trabajo no hubiese sido posible. Gracias por todo lo que me habéis enseñado.

Gracias a Nacho, mi compañero de viaje, y a mis hijas Leire y Elena, por su generosidad dejándome todo el espacio y tiempo necesario que este proyecto ha requerido. Sabré compensároslo.

A Cristina Avendaño Solá, Fernando de Andrés-Trelles, y Gonzalo Calvo (por orden de aparición en mi vida): gracias de todo corazón por descubrirme y enseñarme la parte más excitante de la farmacología clínica. Gracias por ser mis maestros y por vuestra confianza en mí.

Roser Vives y Stella Pesiou, gracias por vuestra inestimable e impagable ayuda en muchos aspectos clave de este trabajo. Stella, gracias por tu generosidad.

A mis compañer@s y amig@s Belén, Isabel, Paloma, Conchi, Gustavo y Rocío, que han estado siempre ahí y no me han dejado tirar la toalla. Paloma, gracias por darle un toque de calidad a todo cuanto tocas, incluido este proyecto.

Y, por último, a mis padres, porque siempre me dejaron seguir mi camino. Sin su esfuerzo y apoyo no estaría donde me encuentro.

BSC	Best supportive care
CHMP	Committee for Human Medicinal Products
COMP	Committee for Orphan Medicinal Products
CAPS	Cryopirine-associated periodic syndromes
CMA	Conditional Marketing Authorisation
EMA	European Medicines Agency
EPARs	European public assessment reports
EC	European Commission
EEA	European Economic Area
EU	European Union
GAS	Goal Attainment Scaling
GEE	Generalised estimated equation
HTA	Health Technology Assessment
ICH	International Conference on Harmonisation
MAA	Marketing authorisation application
MAMS	Multiarm multistage sequential design
MMRM	Mixed effect Model Repeat Measurement
MWS	Muckle-Wells Syndrome
OD	Orphan diseases
ODD	Orphan drug designation
OMP	Orphan medicinal products
PAH	Pulmonary Arterial Hypertension
PEP	Primary endpoint
PIP	Paediatric Investigational Plan
PNH	Paroxismal Nocturnal Haemoglobinuria
QoL	Quality of life
RCT	Randomised controlled trials
RP	Repurposed drugs
SAWP	Scientific advice working party
SOC	Standard of care
UR	Ultrarare
VOD	Veno-oclussive disease
WP	Working packages

1. SUMMARY/RESUMEN

Background

Regulatory uncertainties related to the authorizations of orphan medicinal products (OMP) are often due to unconventional clinical development, hindered by difficulties to recruit patients affected by low prevalence conditions into clinical trials. There is a huge need to strengthen support to developers on the appropriate studies to be conducted in the development of OMP in order to generate adequate evidence for the demonstration of the benefits and risks.

Hypothesis

Alternative methodological approaches to the study of rare or orphan diseases (OD) can be applied to the clinical development of new treatments, which may shorten the period of drug development and reduce the patient needs while keeping integrity and robustness of results.

Guidance to the application of such approaches to clusters of diseases or medical conditions sharing clinical characteristics determining the applicability of critical study design aspects may allow more specific regulatory guidance to researchers aiming to obtain a new marketing authorisation for medicinal products than those currently available.

Objetives

The aims of this project included: 1) to analyse and describe the extent of current regulatory guidance for clinical development of OMP, 2) to select a number of representative examples of authorized OMP suitable for testing of new methods, 3) to qualitatively assess the applicability of a number of novel methodologies and their added value to the design and analysis of clinical trials of new medicinal products aimed for the treatment of OD, 4) to test the applicability of new methods through simulations of drug development plans, and 5) to issue recommendations by groups of conditions that share some characteristics that make them more suitable or adequate for the use of a given methodology.

Methods

The project was developed as a part of a European Collaborative Project funded by the European Commission (EC) (FP7-HEALTH-2013-INNOVATION-1 Grant-Agreement No.603160) with the aim to develop new methodologies for the conduct of clinical trials in small populations. A review of the type of currently available guidance for clinical investigation of OMP, as issued by European Medicines Agency (EMA), has systematized currently available guidance to identify areas where guidance may be further developed. Examples of authorized OMP with available public information in European Public Assessment Reports (EPARs) which were representative of 6 clusterings of conditions were selected. The applicability and the added value of 13 new methodologies and approaches suitable for the study of small populations developed by other working packages within the ASTERIX project was qualitatively assessed, using the framework defined by six condition clusters developed by our group. Applicability testing used real life examples of 26 OMP authorized in the European Union (EU), as published in the EPARs. Further, simulations of alternative clinical development programs were done for 6 selected examples representing the 6 clusters of medical conditions. Regulatory uncertainties were described for each example, as described in EPARs, and these were tried to be solved by

applying alternative approaches. Other relevant methods identified outside of the ASTERIX project could be considered and tested, if deemed appropriate. Based on the outcome of these analyses, an attempt to provide a final set of recommendations as suggestion for implementation in guidelines was done.

Results

Current regulatory guidance specific to clinical development of OMP is scarce and lacks specificity. Applicability of newly developed ASTERIX methods and approaches is limited when just trying to optimize the actually conducted pivotal studies. Simulations of alternative clinical developments applying novel methods revealed that applicability and added value of methods is extended when the aim is set in optimising the drug development program, rather than just improving the pivotal trial as presented in isolation. Novel methods were able to address important regulatory uncertainties and increased the ability of generating robust evidence. The impact of alternative methods and approaches on other relevant ethical or practical aspects varies depending on the methods applied, but when negatively impacted overall this was not to a relevant extent. Based on these results, recommendations on applicability of methods by clusters of medical conditions are proposed.

Conclusions

Existing European regulatory references for clinical development of OMP offer limited and fragmented guidance. Novel methodologies applied to the design and analyses of clinical studies in orphan conditions are useful tools at providing a good balance of robustness and efficiency in the generated evidence and may reduce the level of uncertainties and would facilitate the regulatory decision-making, if applied properly and early in the development planning.

Our framework, methods and recommendations represent an approach to practical and structured thought on the planning of clinical trials and analyses and could support sponsors on the appropriate studies to generate robust evidence in the scenario of small populations.

Introducción

Las incertidumbres reguladoras asociadas a la autorización de comercialización de medicamentos huérfanos se relacionan con frecuencia con la realización de desarrollos clínicos no convencionales, obstaculizados por las dificultades en el reclutamiento de pacientes con enfermedades poco prevalentes en los ensayos clínicos. Existe una necesidad importante de reforzar el apoyo y dar recomendaciones sobre el tipo de estudios a realizar en el desarrollo de medicamentos huérfanos para generar una evidencia adecuada para la demostración de los beneficios y los riesgos de estos medicamentos.

Hipótesis

La aplicación de aproximaciones metodológicas alternativas al desarrollo de nuevos medicamentos en enfermedades raras permitiría que se acorten los periodos de estudio y se reduzca el número de pacientes a incluir manteniendo la integridad y robustez de los resultados. Las recomendaciones sobre la aplicabilidad de estos nuevos métodos pueden hacerse por grupos de enfermedades o condiciones que comparten características clínicas que determinan aspectos críticos del diseño de los estudios, de forma que se podrían elaborar recomendaciones regulatorias más específicas que las existentes que sirvan de guía a los investigadores durante el desarrollo de estos medicamentos.

Objetivos

Los objetivos de este proyecto son: 1) analizar y describir las guías europeas reguladoras existentes aplicables al desarrollo clínico de medicamentos huérfanos, 2) seleccionar un número representativo de ejemplos de medicamentos huérfanos autorizados válidos para evaluar la aplicabilidad de los nuevos métodos, 3) evaluar cuantitativamente la aplicabilidad de un número determinado de nuevos métodos al diseño y análisis de los ensayos clínicos de medicamentos huérfanos y su valor añadido, 4) testar la aplicabilidad de los nuevos métodos a través de simulaciones de planes de desarrollo clínico de medicamentos, y 5) elaborar recomendaciones por grupos de condiciones que comparten características que hacen más adecuada el uso de una metodología determinada.

Métodos

Este proyecto forma parte de un proyecto colaborativo europeo financiado por la Comisión Europea, Proyecto ASTERIX (FP7-HEALTH-2013-INNOVATION-1 Grant-Agreement No.603160) con el objetivo de desarrollar nuevas metodologías para la realización de estudios en poblaciones pequeñas. Se revisaron las guías clínicas europeas aplicables a la investigación con medicamentos huérfanos para identificar potenciales áreas de mejora. Se seleccionaron ejemplos de medicamentos huérfanos con autorización de comercialización europea representativos de los 6 clusters de condiciones definidas previamente en el Proyecto ASTERIX. Se evaluó la aplicabilidad y el valor añadido de 13 métodos nuevos adecuados para el estudio de poblaciones pequeñas desarrollados dentro del Proyecto ASTERIX, utilizando 26 ejemplos representativos de los 6 clusters de condiciones. Posteriormente se identificaron las incertidumbres reguladoras de 6 ejemplos y se simularon desarrollos clínicos alternativos aplicando nuevas metodologías con la intención de resolver las incertidumbres. En base a estos

resultados, se elaboraron una serie de recomendaciones para su potencial implementación en las guías clínicas europeas.

Resultados

Las guías reguladoras existentes para el desarrollo clínico de medicamentos huérfanos son escasas y poco específicas. La aplicabilidad de los métodos de análisis y diseños desarrollados en ASTERIX es limitada cuando se intenta optimizar los estudios pivotales realizados. Las simulaciones realizadas muestran que tanto la aplicabilidad de los métodos como su valor añadido se amplían cuando el objetivo es optimizar el programa de desarrollo clínico del medicamento, no solo el estudio pivotal. Con la aplicación de nuevos métodos se pudieron resolver importantes incertidumbres identificadas durante el desarrollo de medicamentos huérfanos, aumentando así la posibilidad de generar una evidencia robusta. Su impacto en otros aspectos relevantes (éticos o prácticos) fue variable dependiendo del método aplicado, pero en general éstos no se vieron afectados de una forma relevante. En base a estos resultados se hizo una propuesta de recomendaciones sobre la aplicabilidad de los métodos por grupos de condiciones médicas.

Conclusiones

Las recomendaciones reguladoras existentes en la Unión Europea y aplicables al desarrollo clínico de medicamentos huérfanos son limitadas. Los nuevos métodos aplicados al diseño y análisis de estudios en enfermedades huérfanas son herramientas útiles que proporcionan una evidencia con un buen balance entre robustez y eficiencia. Aplicados correctamente y en etapas tempranas del desarrollo pueden reducir el nivel de incertidumbres y facilitar las decisiones reguladoras.

Nuestro marco de trabajo, métodos y recomendaciones representa una aproximación práctica y estructura a la planificación de los ensayos clínicos y análisis y podría ayudar a los promotores sobre los estudios adecuados para generar una evidencia robusta en el escenario de poblaciones pequeñas.

2. INTRODUCTION

2.1. Orphan medicinal products and rare diseases

Orphan drugs are all those medicinal products aimed to diagnosis, prevention or treatment of rare diseases, and are so called because the pharmaceutical industry may have little financial interest in developing and marketing products intended for only a small number of patients suffering from very rare conditions (1).

In the EU, rare diseases are the medical conditions which have a prevalence figure of less than 1 person per 2,000 inhabitants (2). A disease can be rare in one region, but common in another. This is the case of thalassemia, an anemia of genetic origin, which is rare in Northern Europe, but it is frequent in the Mediterranean region. This prevalence figure is arbitrary and established to determine the threshold below which the market conditions are no longer conventional, and special needs are displayed for drug development. In fact, the definition of a rare disease is not universal and depends on the legislation and policies adopted by each region or country (3). In this sense, the definition of a rare condition in the US is wider than in the EU (prevalence approximately 7.5 in 10,000 inhabitants), while in Japan is stricter (approximately 3.9 in 10,000 inhabitants) and even more in Australia (approximately 1 in 10,000 inhabitants) (3).

The estimated overall numbers of existing rare diseases range from 6,000 to 7,000 distinct diseases. Despite being infrequent, the high number of conditions makes rare diseases anything but rare, affecting between 6% and 8% of the population in total, which means between around 27 million and 36 million people in the EU. Most people suffer from diseases affecting fewer than 1 in 100,000 people (4).

In up to 80% of rare diseases genetic origins have been identified and affect between 3% and 4% of births. Other rare diseases are the result of infections (bacterial or viral), allergies and environmental causes, or are degenerative and proliferative. To date, the cause remains unknown for many rare diseases (2).

Rare diseases often are serious, life-threatening or chronically debilitating and progressive diseases. They decrease the quality of life and performance status of patients affected, and generally causes long-term disabilities and dependence. Rare diseases not only affect the person diagnosed - they also impact families, friends, care takers and society as a whole (5).

For many rare diseases, signs may be observed at birth or in childhood. Examples include proximal spinal muscular atrophy, neurofibromatosis, osteogenesis imperfecta, and chondrodysplasia or Rett syndrome. However, over 50% of rare diseases appear during adulthood, such as Huntington diseases, Charcot-Marie-Tooth disease, amyotrophic lateral sclerosis, Kaposi's sarcoma, thyroid cancer or acute myeloid leukaemia (6).

Subjects affected by these diseases all face similar difficulties in their quest for a diagnosis, relevant information and proper direction towards qualified professionals. Specific issues are equally raised regarding access to quality health care, overall social and medical support, effective liaison between hospitals and general practices, as well as professional and social integration and independence (3).

Fortunately, hundreds of rare diseases can now be diagnosed through a biological sample test. Knowledge of the natural history of these diseases is improved by the creation of registries. Researchers are increasingly working through networks in order to share the results of their research and to advance more efficiently. New hopes arise with the perspectives offered by dedicated policies in the field of rare diseases, but still there is lot of work to be done (2).

2.2. Peculiarities of orphan drug development

Conducting a pharmaceutical development program for the treatment of a rare disease is particularly challenging. Among the main difficulties are the poor understanding of the natural history of the proposed indication, due to lack or few observational studies on disease progression, difficulties to achieve an accurate diagnosis, heterogeneity of patient populations with variable phenotypes and clinical courses, limited number of patients plus geographic dispersion of patients and investigators, regulatory uncertainties, difficulties in choosing clinically relevant outcomes, and lack of prior experience in conducting clinical studies that could serve as guidance (7).

Limited knowledge on natural history of rare diseases

Unlike more prevalent diseases, there is little existing knowledge on the natural history of most rare diseases. Studies designed to describe natural history can yield vital information for the design and conduct of clinical studies for product developments. These studies can provide valuable information about demographic, genetic, and environmental variables that correlate with the time course, stages and outcome of the disease; refinement of diagnostic criteria and identification of patient subpopulations with different characteristics and effects of the disease; patient perspectives on what aspects of disease are most important to treat; and how to quantify those aspects, so that they can serve as useful outcome measures for clinical trials. In the case of rare diseases, their natural histories frequently are not fully understood because there are simply not enough diagnosed cases that have been observed and studied (8).

This lack of knowledge limits researchers' ability to study rare diseases and develop new treatments. Knowledge of natural history is essential for developing more efficient clinical trial designs. It also could help reduce the length and cost of drug development and, possibly, contribute toward greater predictability of clinical development programs (8).

Researchers are faced with limited approaches to appropriately address these challenges. Patient registries can help to fill this gap due to their value in developing evidence for rare diseases. There are some significant operational challenges that need to be considered and addressed in managing these programs. Currently, a number of international efforts seek to promote, standardize and facilitate these efforts and will have a strong impact on improving evidence development and patient care (8).

Underdiagnose

Patients with rare or complex, life-threatening diseases often face challenges in obtaining a correct diagnosis and in timely accessing appropriate therapies and clinical expertise. The small prevalence of the conditions means that in general their physicians have never seen similar cases and are thus less able to suspect and establish a precise diagnosis. Often patients are left untreated or have to scour the internet in the hope of finding a center with the necessary expertise. The high number of conditions and their complexity makes it difficult to ensure that there is enough awareness on each independent disease by clinicians, and because of the same reason, there is limited specific disease expertise, which may be geographically concentrated and at significantly distance from patients as to physically impair access to healthcare. In order to improve this situation, the European Reference Networks (ERN) initiative was set in order to connect and ease exchange of information and expertise on rare diseases across the European region. Such networking may also ease clinical trials by connecting centers of expertise at the national and international level (9).

Further, there are difficulties in identifying participants for clinical research due to lack of proper identification of patients with many OD in health information systems. Scarcity of cases makes it difficult to establish clear diagnosis criteria, because of lack of series large enough as to inferring common features that are determinant to establish a medical diagnosis. In addition, disease coding and recording in medical records may be heterogeneous and lack clinical specificity. Most countries use the International Classification of Diseases (ICD-10) to record patients, where around 500 rare diseases have a specific code. In countries using Systematized Nomenclature of Medicine (SNOMED), the situation is not much better because only around 40% of rare diseases are listed here. Thus, any attempt to establish disease prevalence and to identify patients in a given region represents a challenge at the medical coding and recording.
Another source for identification of rare disease patients is disease-specific patient registries. There are 690 such registries in Europe, covering 984 rare diseases. Most are national (482 registries), or regional (75 registries), with some being European (59 registries) or international (74 registries). However, quality, scope, and capacity of many registries are limited (10).

Other options to connect patients and increase feasibility of research in sized-enough populations include patient networking and self-organization. EURORDIS is a non-governmental patient-driven alliance of patient organizations representing 808 rare disease patient organizations in 70 countries, giving voice to 30 million people affected by rare diseases throughout Europe. Patient organizations represent a key stakeholder to increase awareness on active research and to connect researchers with patients, easing access to recruitment into clinical trials (11,12). Similarly, recently patient groups from all over Europe have come together to launch a new coalition, Syndromes Without A Name (SWAN) Europe, uniting European and national patient organisations to support and empower families affected by syndromes without a name. The initiative is focused on providing a forum for sharing information and providing a point of contact for stakeholder engagement across Europe. Such a patient-driven initiative may serve as a platform to increase feasibility of better diagnosis and access to relevant clinical research (13).

Small and heterogeneous patient groups

The randomized controlled trial (RCT) is considered the gold standard for establishing efficacy in a research setting. This design literally implements the scientific method by prospectively comparing the effects of a given intervention as the only difference between two a priori identical groups, allowing concluding on potential causality of a given treatment on patient outcomes. Through randomization minimizes selection bias and equally distributes potential confounders, known and unknown, between study groups. Blinding minimizes the potential for confusion or bias due to patient and investigator expectations. Together, randomization and blinding have the potential to limit investigator and participant bias in outcomes assessment. Use of controls will strengthen trial design by addressing concerns regarding clinical variability (14).

Despite there is consensus on the fact that randomized clinical trials are the gold standard to obtain evidence on efficacy and safety of new treatments, several authors have found that the pivotal studies for orphan drug approvals were more likely to be smaller, use non-randomized, unblinded trial designs, and less likely to use placebo control. Most pivotal studies for orphan drugs are single arm designs and use surrogate endpoints to assess efficacy, and only rarely apply alternative trial designs that may be appropriate to methodologically manage uncertainty in exceptional situations with limited sample size: a survey in ClinicalTrials.gov showed that Bayesian methods and adaptive randomization, although recommended in the relevant EU guidelines are uncommonly used (15–17).

The classic clinical trial design, almost universally used to detect small therapeutic effects for treatments aimed to common diseases, requires large sample sizes and consequently is costly and time-consuming. Large samples are less feasible in rare diseases, and controlled designs and randomization can prove to be difficult with rare diseases when a reasonable therapeutic alternative does not exist (15–17).

In rare diseases, many of which are paediatric and cause a shortened lifespan (18), there are ethical concerns about placebo-controlled trials. Parents may be reluctant to enroll their child in a trial where he or she may receive a placebo rather than the intervention under study. Balanced against the clinical researcher's desire to maintain equipoise is a likely assumption by hopeful families of an expected clinical benefit, regardless of the fact that this is yet to be proven. In a rapidly progressing fatal disease, there is perhaps greater urgency on the part of parents to ensure their child is exposed to an active treatment condition, before the possible window of therapeutic opportunity is lost (19).

By necessity, clinical trials in rare disorders enroll small samples. In combination with high inter-individual variability in clinical course observed in many rare diseases, this diminishes a study's power. Consequences of suboptimal designs include impaired robustness of the evidence on clinical efficacy and safety, thus increasing the degree of inference applied to decision making, and the degree of empiricism in clinical practice. Besides, the assumption of efficacy based on weak evidence may nevertheless establish new standards of care, reducing the perceived need or willingness to develop better approaches, or defining reference treatments against whose any future therapy may compare with non-inferiority approaches. If such

comparisons are done to, in example, test marginal improvements such as more convenient posology, this may perpetuate uncertain or non-effective therapies for the disease (20).

Because of that, while the complexity of performing clinical trials in orphan populations should be acknowledged, methodological designs should still strive to include blinding and randomization, which are among the hallmarks of high-quality clinical trial design (19–21).

Recruitment issues due to rarity and patient dispersion

Smaller prevalence and geographic dispersion imply difficulties in identifying and locating potential participant patients. Rare diseases affect very few people, and thus patients are often widely dispersed in geography, and are heterogeneous in disease subtype, symptoms, stages and exposure to prior treatment. This adds difficulties in finding enough patients who fit inclusion and exclusion criteria for a particular trial. Wide geographic dispersion may require either finding enough clinical sites with the required expertise or implementing measures for patient transport to clinical sites for assessment, inclusion and follow-up visits. Further, this would require developing research documents and product labeling in several languages, going through complex international regulatory procedures for trial authorization, and complex logistic arrangements, all of which may complicate trial protocol development and administration, data collection, outcomes measurement, thus increasing exponentially development costs.

When studying an orphan disease, every single patient's participation is vitally important given limitations in patient availability, and the exceptional impact the data from a limited number of patients may have on program development. In that sense, eligibility criteria always influence the number of available subjects, and if artificially constrained in order to control heterogeneity and preserve the trial power, this would make recruitment even more difficult and reduce the likelihood of achieving database large enough as to obtain evidence of efficacy and safety. Strict inclusion criteria may also compromise external validity, so that evidences from a small but homogeneous trial may not be extrapolated to a larger network of representative patients with the same disorder (10).

Other regulatory uncertainties of small trials

For orphan drugs, the usual stepwise development in phases might not be so well defined and approval may be granted without a typical clinical trial program. Though it is specified that the minimum exposure requested for conventional drugs according to international regulations (International Conference on Harmonisation (ICH, E1A) does not always apply to orphan drugs, it is not defined what is expected. The size of a clinical database as defined by the ICH E1 is expected to characterise and quantify the safety profile of a drug over a reasonable duration of time, consistent with the intended long-term use of the drug, so that short-term event rates (cumulative 3-month incidence of about 1%) and rare adverse events occurring in more than 1 in 1,000 patients(22).Thus, not meeting the intended sample size as defined by ICH E1 means that clinical safety evidence may deal with greater uncertainty in rare diseases than in conventional and more prevalent diseases, in contradiction to the official statement in the European Regulation on OMP *"patients suffering from rare conditions should be entitled to the same quality of treatment as other patients"* (23).

Small prevalence is a solid argument against following the rule on size of the product clinical database, since lack of enough affected patients prevents to reach the relevant numbers, but while such an argument allows flexibility for individual features of unique drug applications based on scientific judgment, otherwise it causes uncertainty at the time of establishing an appropriate risk/benefit assessment. Further, this leaves drug manufacturers without guidance on what to provide with the approval package, particularly when not prior experience does exist on relevant aspects of the study design as relevant as the selection of a relevant primary endpoint (PEP) for the demonstration of efficacy, increasing investing risk (24). The so called procedure of Scientific Advice, offered by the EMA, may in part cover the lack of regular guidance for a given clinical situation, improving the predictability of future regulatory requirements and acceptability of particular approaches (25,26).

Early regulatory advice may reduce the development risk for the industry, but though useful, those scientific recommendations given in advance are not binding to the final regulatory reviewer (25). Lack of binding policy regarding specific regulatory requirements for approval of orphan drugs is a risk perceived both by the industry and advocacy groups (27).

All these challenges account for the high cost of orphan drugs developments. The whole process to complete the discovery and development process for any type of drug could last an average of 12-13 years. A big capital investment is needed and only a few projects will finally succeed. The process is characterized by high investment requirements and high risk of failure, both before reaching the market and even after the marketing approval. For OMP this should be considered in the context of a low number of patients affected, which in addition to the previous limitation supposes a reduced market size for an OMP, limiting the potential for return of investment. In that sense, in the EU the maximum market size for a particular rare disease would be 257,850 patients to be treated; this number was taken into account as an estimation of potential lack of commercial interest at the time of the definition of the EU threshold to consider that a given treatment is deemed an Orphan Drug Designation (ODD)(23,28).

These particularities explain why orphan drugs development pose unique challenges, mostly due to the logistical difficulties of working with small patient population widely geographically dispersed, added to the fact that often OD are serious and debilitating conditions with limited or not at all treatment options available. In spite of the advances made during the past years, and a growing interest of pharmaceutical companies in the increasingly profitable field of rare diseases, recent analyses conclude that difficulties in orphan drug development persist (29).

2.3 The legal framework of orphan medicinal products

As commented above, drug development for rare diseases is often limited by the prohibitive cost of investing in an original pharmaceutical agent as compared to the limited profit potential given by the small patient size per rare disease indication. Under human rights principles, the EU regulation recognizes that patients with rare diseases have equal rights to medicines as other patients with more prevalent disease (e.g. diabetes). They should not be excluded from gaining benefits from medical advances just because of the rarity of their illness, nor should be exposed to treatments with lower standards of quality, efficacy nor safety (23).

In this context, many governments and authorities have established legislations, regulations and policies to encourage the research and development of orphan drugs and to address licensing regulations and pricing and reimbursement of these drugs; such economic and regulatory incentives are important public health decisions (30–32).

The concept of the orphan drug was introduced in the USA in 1983. The US Orphan Drug Act encourages investment in rare disease research by giving a product an additional period of seven years of market exclusivity, as well as tax credits and other incentives for any money spent on this type of research. To qualify as an orphan drug, the disease must affect fewer than 200,000 Americans. A similar legislation to the one of US did not appear in Europe until 2000 when the 'orphan medicinal product' category was introduced (33).

The first EU legislation related to orphan drugs was Regulation (EC) No 141/2000 (the Orphan Regulation) (23), approved on 16th December 1999 and entered into force on 22 January 2000. This regulation establishes 3 main aspects:

1. *It lays down the EU procedure for designation of orphan medicines*. To qualify for orphan designation, a medicine must meet the following criteria:

    a) it must be intended for the treatment, prevention or diagnosis of a disease that is life-threatening or chronically debilitating;

    b) the prevalence of the condition in the EU must not be more than 5 in 10,000, or it must be unlikely that marketing of the medicine would generate sufficient returns to justify the investment needed for its development;

    c) also no satisfactory method of diagnosis, prevention or treatment of the condition concerned can be previously authorized, or, if such a method exists, the medicine must be of significant benefit to those affected by the condition.

2. *It defines incentives for the development and placing onto the market of designated orphan medicines*. This includes a special form of scientific advice for orphan drugs called Protocol Assistance, available at a reduced charge for designated orphan medicines, that depends on the status of the sponsor, access to the centralized authorization procedure, additional incentives for small, medium-size enterprises, fee reductions, priority to access grants and additional incentives from Member States, as well as the most attractive incentive, i.e. the 10-year market exclusivity protecting against competition from similar medicinal products, which may be extended 2 additional years if data on the paediatric population according to an agreed Paediatric Investigational Plan (PIP) are provided.

3. *It establishes a new Committee within the EMA, the Committee for Orphan Medicinal Products (COMP)*, responsible for evaluation applications for ODD.

Subsequently, Commission Regulation (EC) No847/2000 of 27 April 2000 (34)was approved to establish provisions for implementation of some articles included in the Regulation141/2000 (specifically articles 3 and 8). The document assists all stakeholders in the interpretation of the following aspects of the EU Regulation:

- Criteria for designation of a medicinal product as an OMP. These consider the prevalence of the medical condition treated, the potential return on investment, the description of methods of diagnosis, prevention and treatment, and general provisions.

- Definitions of the concepts “significant benefit”, “similar medicinal product” and “clinical superiority”.

After this date the sponsors began to submit the applications to obtain an ODD.

These two regulations are the major legal basis to foster impulse on research and development of medicines for rare diseases in the EU. Additional legislations have been approved later on, which have different degrees of implications on the regulation of OMP. Regulation (EC) No 726/2004(25)determines that all marketing authorisations for orphan medicines in the EU should follow the centralized authorisation procedure. This implies a single marketing authorisation to EMA, a single scientific assessment and recommendation carried on by the Committee for Human Medicinal Products (CHMP) of the EMA, and finally a single marketing authorisation granted by the EC valid in all EU member states and the European Economic Area (EEA) countries. Regulation (EC) No 507/2006(35), approved on March 2006,provides the legal framework for the granting of a conditional marketing authorisation (CMA) to medicines that fall within the scope of Regulation (EC) No 726/2004. It establishes that orphan medicines can be granted a CMA within this legal framework. Regulation (EC) No 1901/2006 (36)approved on December 2006, also known as paediatric regulation, included a 2-year extension of the market exclusivity period for orphan medicines which fulfill the requirements for generation of data regarding their potential use in paediatric population. Finally, Regulation (EC) No 2049/2005(37) approved on December 2005, established that the Scientific Advice procedure requested by small, medium-size enterprises for OMP designated by the COMP would be fully free of charge (37).

All these legislations have entered into force during the last 18 years and constitute the legal framework adopted in the EU to promote and facilitate the development of OMP. Published reviews have shown that an important portion of orphan medicines are developed by small to medium-sized enterprises, explaining why these incentives have generally been regarded as being successful in stimulating the development of medicines for rare diseases (38).

At present, many countries in the world have in place their own regulations and policies for orphan drugs. Because rare diseases are a global issue, collaboration among international regulatory partners on the designation and assessment of orphan medicines, is critical to facilitate orphan drug developments.

2.4 Marketing authorization of orphan medicinal products in Europe

2.4.1 Evaluation process

The regulatory process for OMP starts with an ODD application. The ODD application is evaluated by the COMP. To obtain and maintain ODD, specific criteria need to be fulfilled, such as the life-threatening and/or seriously debilitating nature of the disease and a demonstration of its prevalence. In addition, the proposed medicine should be medically plausible and there should either be no available treatment, or the new medicine should be able to demonstrate significant benefit over existing treatment options. Significant benefit is defined in Regulation (EC) 847/2000 (34) as a clinically relevant advantage or a major contribution to patient care. The ODD is granted on preliminary data that support an assumption of significant benefit. An ODD does not mean nor assure a marketing authorisation, but the Company can benefit from incentives such as scientific advice on study protocols, various fee reductions and access to EU grants.

Orphan-designated medicines that reach the marketing authorisation application (MAA) stage are evaluated by EMA's CHMP using the same strict quality, safety and efficacy standards that apply to all medicines evaluated by the EU regulatory network. Following the scientific assessment of the application, the CHMP will determine whether the medicine meets the necessary quality, safety and efficacy requirements and that it has a positive benefit-risk balance. Based on the outcome of this assessment, the CHMP gives a recommendation to the EC on whether the medicine should be marketed or not. A marketing authorisation granted by the EC is valid in all EU Member States as well as in the EEA countries Iceland, Liechtenstein and Norway.

In accordance to the Orphan Regulation (EC) No 141/2000(23), a marketing authorisation for any medicinal product (orphan or non-orphan) can only be granted provided that no similar OMP are currently under market exclusivity protection for the same therapeutic indication. If this were the case, the Applicant will need to submit a similarity report addressing the possible similarity between the new medicinal product and the OMP(s) under market protection. If based on the mechanism of action, and/or the molecular structure and/or the therapeutic indication, the new medicinal product were considered similar to any of the marketed products under market protection, the CHMP could only recommend the marketing authorisation of the new product provided that one of the derogations provided for in Article 8(3) of the Orphan Regulation (EC) No 141/1200 claimed by the applicant applies (i.e. the holder of the marketing authorisation for the original OMP has given his consent to the second applicant, or the holder of the marketing authorisation for the original OMP is unable to supply sufficient quantities of the medicinal product, or the second applicant can establish in the application that the second medicinal product, although similar to the OMP already authorized, is safer, more effective or otherwise clinically superior). Otherwise, the CHMP will reject the MAA regardless of the quality, safety and efficacy data of the new medicinal product.

In parallel, the COMP will re-evaluate whether the medicine with a CHMP positive opinion continues to meet the criteria for maintaining its orphan status. The sponsor will need to submit a report of *maintenance of the orphan designation* in parallel to the MAA (or extension of an existing marketing authorisation), including data on the current prevalence of the condition to

be diagnosed, prevented or treated, or the potential return on investment, the current life-threatening or debilitating nature of the condition, the current existence of other methods for the diagnosis, prevention or treatment of the condition and, if applicable, a justification of the medicine's significant benefit. Based on the data available at the time and on the sponsor's report, the COMP will determine whether the medicine can maintain its status as an orphan medicine and benefit from market exclusivity (39).

Orphan-designated medicines that eventually make it to the market, and for whom it can be demonstrated that they maintain the criteria for the designation, are granted 10 years of market protection.

The spirit of the orphan legislation is to promote the development of novel therapies for rare conditions. Data recently published by the EMA concerning ODD and orphan drugs marketing authorisation figures since 2000 up to the end of 2017 support the success of the EU policies. Indeed, the number of designations has continuously increased since 2000 and, as of the end of 2017, more than 1,900 designations had been granted by the EC. Approximately 40% of designations were for conditions with a prevalence of less than 1 in 10,000 people in the EU, with 13% of the designations for purely paediatric diseases (40).

By the end of 2017, 142 orphan medicines had been authorized in the EU (plus 20 extensions of the indication of previously authorized medicinal products) that could be considered to have benefitted from the orphan incentives. This includes withdrawals from the register of OMP, the register of medicinal products for human use, and expired orphan status. These total 162authorisations represent a total of 111 different rare conditions with more than 40% being authorised in the area of oncology, followed by alimentary tract and metabolic diseases (19%). As with the designations, almost half (49%) of the authorized products are for diseases with a prevalence of less than 1 in 10,000 (40).

2.4.2. Marketing authorization under special conditions

The EU Regulation has provisions for the early access to medicines that address unmet medical needs of patients. In the interest of public health, applicants may be granted a “special” marketing authorisation for these medicines based on less comprehensive data than normally required where the benefit of immediate availability outweighs the risk of this limited evidence. These tools are the conditional marketing authorisation and the authorisation under exceptional circumstances (41).

Conditional marketing authorisation

The CMA is established in Regulation (EC) No 507/2006 and is intended for medicines that address an unmet medical need, and that target seriously debilitating or life-threatening diseases, rare diseases, or is intended for use in emergency situations in response to a public health threat. This tool allows for the early approval of a medicine on the basis of less complete clinical data than normally required. However, these medicines are subject to specific post-authorisation obligations that aim to obtain complete data on the medicine to, ultimately, recommend a full marketing authorisation if the complete data confirms that the benefits of the medicine outweigh the risks.

A CMA may be granted if the CHMP finds that all the following requirements are met:

- The benefit-risk balance of the product is positive.
- It is likely that the applicant will be able to provide comprehensive data.
- Unmet medical needs will be fulfilled.
- The benefit to public health of the medicinal product's immediate availability on the market outweighs the risks due to need for further data.

CMA is valid for one year and can be renewed annually. Once comprehensive data on the product have been obtained, the marketing authorisation may be converted into a standard marketing authorisation (not subject to specific obligations). Initially, this is valid for 5 years, but can be renewed for unlimited validity.

The granting of a CMA will allow medicines to reach patients with unmet medical needs earlier, while ensuring that additional data on a product are generated, submitted, assessed and acted upon.

Authorisation under exceptional circumstances

A marketing authorisation in absence of comprehensive data may also be granted under exceptional circumstances. Unlike CMA, where marketing approval is granted in the likelihood that the sponsor will provide such data within an agreed timeframe, authorisation under exceptional circumstances can be granted when comprehensive data cannot be obtained even after authorisation. This authorisation route normally does not lead to a standard marketing authorisation.

The legal basis for the marketing authorisation under exceptional circumstances is the Article 14 (8) of the Regulation (EC) No 726/2004, and the relevant documentation for applications in exceptional circumstances are laid down in Part II of Annex I of Directive 2001/83/EC, as amended (25,35). It is only valid for medicinal products for which the applicant can demonstrate in this application that he is unable to provide comprehensive data on the efficacy and safety under normal conditions of use, because:

- The indications for which the product in question is intended are encountered so rarely that the applicant cannot reasonably be expected to provide comprehensive evidence,

or

- in the present state of scientific knowledge, comprehensive information cannot be provided,

or

- it would be contrary to generally accepted principles of medical ethics to collect such information.

Consequently, the authorisation under exceptional circumstances is granted subject to a requirement for the applicant to introduce specific procedures, in particular concerning the safety of the medicinal product, notification to the competent authorities of any incident relating to its use, and risk minimization actions to be taken.

2.5. Regulatory uncertainties of the approval of OMP in EU

Scientific evidence that support the MAA for an orphan drug has in general lower quality than that of non-orphan medicines (15–17). The lower the robustness of the evidence the higher the level of uncertainties with which a decision must be taken.

2.5.1. Heterogeneity in assessment criteria

In the absence of specific EU scientific guidelines for the clinical evaluation of new medicinal products for the treatment of the vast majority of rare conditions, often less than comprehensive data and/or lower quality than normally required are submitted as the main support for a MAA. The limited experience from regulators when dealing with new unexplored fields, considerations on unmet needs and feeling of emergency to offer solutions to the medical condition and the relative methodological weakness and inferential value of the data, all together may explain the huge heterogeneity in the interpretation of the available evidence across the different decision-makers. The robustness and validity of data are key factors for the interpretation of the evidence, so that the lower the robustness of the evidence, the higher the level of uncertainties and subjectivity in the assessment and decision making. If the level of uncertainties is too high, this may end up in the rejection of the marketing authorisation due to difficulties in making a proper benefit/risk balance. In less extreme cases, a positive benefit/ risk balance might be concluded, but the relevant uncertainties identified will need to be addressed during the post-marketing, which implies a huge burden for companies and regulators. Thus, reducing the uncertainties may facilitate the decision-making process for OMP at different levels in the access to market.

2.5.2. Amount and quality of data

To assess the level of uncertainty in regulatory decision-making for OMP a revision of the current basis for approval is needed.

The type of evidence and characteristics of the clinical trials that support marketing authorisations for OMP have been recently described in a systematic review conducted by our group (42). The revision included a total of 125 dossiers from authorized OMP published between 1999 and 2014 on the EMA website and provides an insight into the current regulatory standard. In brief, the results showed that up to 88% (110/125) of OMP authorisations were based on clinical trials, with 35% (38/110) including replicated pivotal trials. The mean (SD) number of pivotal trials per indication was 1.4 (0.7), and the EPARs included a median of three

additional non-pivotal supportive studies. Ten percent of EPARs (13/125) were authorized in spite of negative outcomes from pivotal trials. One-third of trials (53/159) did not include a control arm, one-third (50/159) did not use randomization, half the trials (75/159) were open-label, and 75% (119/159) used intermediate or surrogate variables as the main outcome. The study concluded that regulatory evidence supporting OMP authorisation showed substantial uncertainties, including:

- Weak protection against errors, due to small and inadequately dimensioned and powered trials that increases the risk of falsely concluding on relevant efficacy.

- Substantial use of designs unsuited for obtaining conclusions on causality, mainly due to frequent use of uncontrolled designs that are susceptible to many potential biases and confounding, and thus also increasing the risk of overestimation of treatment efficacy.

- Use of intermediate variables, lacking proven translation of any observed effects into relevant effects on clinical outcomes, and increasing again the risk of efficacy overestimation.

- Lack of apriorism, with new approaches to study analysis that may be data driven, thus increasing the risk of bias and highlighting of chance findings that may lead to overestimation of effects.

- Insufficient safety data, impairing the ability to quantify risks of relevant magnitude.

The results of this revision were consistent with previous findings and highlight the existence of numerous and important areas of uncertainty at the time of regulatory decision-making on OMP(43).

2.6. Regulatory actions aimed to improve orphan drug development

Besides the inherent difficulties encountered when conducting clinical trials in small populations, there is data indicating that the pivotal studies which are considered the basis for the OMP marketing authorisation, exhibit sometimes methodological flaws, consequently leading to a need for more demanding regulatory process and assistance to the sponsors (43).

In order to increase the success of the development (24), and thus the availability of treatments for rare diseases, the EU regulatory network provides advice to sponsors. Advice to product development can be given on a case by case basis, through the scientific advice procedure. Alternatively, general advice for frequently asked questions or aspects on product development can also be given through the issue of scientific guidelines (26).

2.6.1. Scientific advice and protocol assistance

Scientific advice is a regulatory procedure where the EU Regulatory Agencies (EMA or National Agencies) give advice to a developer on the appropriate tests and studies to be conducted for the development of a medicine. At EMA level, scientific advice is given by the Committee for Medicinal Products for Human Use (CHMP) on the recommendation of the Scientific Advice Working Party (SAWP) (26).

Scientific advice

Scientific advice helps to ensure that developers perform the appropriate tests and studies, so that no major objections regarding the design of the tests are likely to be raised during evaluation of the MAA. Major objections can significantly delay the marketing of a product, and, in certain cases, may result in refusal of the marketing authorisation. So, scientific advice is a regulatory initiative aimed to reduce the risk of failed developments by supporting development programs and early dialogue. Following the Agency's advice increases the probability of a positive outcome (26,40).

The advice is given in the light of the current scientific knowledge, based on the documentation provided by the medicine developer. Scientific advice is prospective in nature. It focuses on development strategies rather than pre-evaluation of data to support a MAA. Developers of orphan medicines can receive answers to questions related to the criteria for authorisation of an orphan medicine. These include:

- The demonstration of significant benefit within the scope of the designated orphan indication.

- Similarity or clinical superiority over other medicines. This is relevant if other OMP exist that might be similar to the product concerned and which have market exclusivity in the same indication.

Scientific advice received from the Agency is not legally binding on the Agency or on the medicine developer with regard to any future MAA for the medicine concerned (44).

Protocol assistance

One of the incentives offered in the EU for medicines that have been granted an orphan designation by the EC is the so called *Protocol assistance*, a form of scientific advice for orphan medicines. This allows sponsors to get answers to their questions on the design of studies needed to demonstrate the medicine's quality, benefits and risks, identical to the normal scientific advice.

Protocol assistance is available at a reduced charge for designated orphan medicines, linked to a fee-reduction scale that depends on the status of the sponsor. There is no restriction on the number of times a sponsor can request protocol assistance.

Parallel EMA-FDA advice

Because rare diseases are a global issue, collaboration among international regulatory partners on the designation and the design of development programs for orphan medicines, is critical to facilitate orphan drug developments. To this aim, there exists the possibility to seek parallel scientific advice from the two main regulatory agencies (EMA and FDA), so that joint discussions with both regulatory bodies at an early stage of development may allow to align requirements and to generate evidence able to satisfy the requirements from both regulatory bodies (45).

Parallel consultation with regulators and health technology assessment (HTA) bodies

In addition, the CHPM-SAWP offers consultations in parallel with European Network for HTA Agencies (EUnetHTA) (46). While the main focus of the regulatory assessment is to obtain evidence enough on the benefits and risks of the intended use of the drug to support the marketing authorisation decision, the perspective of HTA is to obtain information enough on the therapeutic utility or the added value as to define the ideal therapeutic positioning of the innovative product and its economic value, in relation to already existing alternatives. The different perspectives often create difficult scenarios where a drug may be authorised for a given use, but the supportive data may not be appropriate to demonstrate added value of the new treatment, making patient's access to the treatment difficult or not feasible (47). In Europe, the parallel EMA-HTA advice procedure allows an early interaction with sponsors, such that OMP developers can obtain feedback from regulators and HTA bodies on their development plans and their ability to generate the needed evidence for the regulatory and national health systems decision-making of new medicines at the very same time (45).

The importance of making use of this regulatory incentive is illustrated by the analysis conducted by Matthias P. Hofer at al. (40), on the factors determinant for orphan marketing authorisation outcome. The analysis was conducted on the experience gained with orphan marketing authorisation from the first 14 years of the orphan EU regulation. The data showed that there was a higher likelihood for MAA success when sponsors received and complied with recommendations on the clinical development (in particular, primary efficacy endpoint, comparator, and statistical methodology) compared with sponsors that were noncompliant with scientific advice recommendations (80% versus 36%). The other factor was the company size; MAA submitted by small companies were found to be less successful regarding the outcome of CHMP evaluation compared with MAA submitted by medium-sized and large companies (small: 54%, medium: 76%, large: 79%). This could be explained by the lower regulatory experience, less use of the EU incentives like scientific advice/protocol assistance, and a high trend to solve the challenges faced during orphan drugs development with the use of smaller clinical development programs than those used by large companies.

The strong association between compliance with scientific advice recommendations on clinical trial design and marketing authorisation success for orphan drugs confirms the importance for developers of orphan medicines to take advantage of the EU orphan incentive system and engage in early regulatory dialogue.

2.6.2. Scientific guidelines

The EMA's CHMP prepares scientific guidelines in consultation with regulatory authorities in the EU Member States, to help applicants prepare MAA for human medicines. These guidelines reflect a harmonized position of the EU Member States and the EMA on how to interpret and apply the requirements for the demonstration of quality, safety and efficacy set out in the Community directives (48).

Several scientific guidelines with recommendations to support applicants and/or sponsors with the overall pharmaceutical product development, as well as the nonclinical and clinical studies of a product intended to be used in humans, have been redacted over the years within the EU Regulatory Network. Those are related to specific scientific issues reflecting a harmonized EU approach on the investigation and are based on the most up-to-date scientific knowledge (48). Adopted guidelines are subjected to a periodic maintenance and revision based on relevant changes in scientific knowledge and/or regulatory experience.

In general, the relevant parties (i.e. applicants, sponsors, manufacturers) must take under consideration and follow these scientific guidelines for the preparation of their applications for marketing authorisation; otherwise, deviations from the guidelines must be appropriately justified. Guidelines are also useful for assessors during the evaluation of MAA. Adherence to scientific guideline recommendations will facilitate the procedures of assessment, approval and control of medicinal products in EU.

Among the guidelines that have come into effect in the EU until now, several of them relate to the status of rare diseases and OMP, e.g. the guideline on clinical trials in small populations (49); however, many gaps in guidance regarding this research field are still responsible for a high number of questions addressed to EMA. Seek of scientific advice for topics like the limitations imposed by the rarity of the disease and the collection of comprehensive data on safety and efficacy, or the acceptability of alternative designs remains high when a potential OMP is under investigation (50).

Therefore, it is important to ensure that appropriate advice is available for the development of medicinal products for orphan conditions through EMA guidelines. Nevertheless, the actual extent and availability of EMA guidance on this topic, and whether there is still room for improvement has not been systematically reviewed.

2.7. Need for new methodological approaches for clinical trials

Randomized controlled trials (RCT) are considered the reference standard in clinical research. It is a solid, intuitive and robust design that allows drawing consistent and reliable conclusions on causality, provided that the study is conducted in accordance to a previously authorized and well-designed study protocol (14). RCT is the preferred design for regulators also because it allows minimizing the risks for the public health that a high type I error might have. Thus, it is also the preferred study design for companies during clinical drug development as long as it satisfies regulatory requirements (51).

However, RCT are criticized for being costly (in terms of number of patients required), long-lasting (given that patients should stay until all patients complete the study), inflexible (due to the fact that design parameters, often chosen with uncertainty, are not rechecked until the end of the trial, and also because the information is no used as it is gathered, but only at the end of the trial), and finally, it constitutes a stand-alone evidence (as long as prior available information is only used for sample size calculation but otherwise disregarded).

These limitations make the conduct of large RCT rather challenging, particularly when dealing with small populations, like those affected by OD and children. However, even in these situations, the EU guideline on small populations clearly states that limited evidence from a small high quality clinical trial is better than no data at all (20,49).

The prior limitations of small clinical trials can be addressed by making the best use of available patients throughout measures aimed to maximize recruitment, keeping long trials simple to reduce drop outs, identifying components of variance and minimizing them, measuring all relevant endpoints, and optimizing study designs. When measures aimed to increase the efficiency of clinical trials are not enough, and controlled evidence is not feasible to be obtained, the regulatory assessment may accept approaches that relax statistical requirements like e.g. relaxing the probability of type I error or by using historical controls, provided that patients' interests are duly protected.

There are different ways to optimize study designs without compromising the generation of reliable, clinically relevant interpretable results. All of them rely on the premise that RCT, even if severely underpowered, are better than series of cases.

They might be classified into 4 categories based on the main consequence of its implementation:

- Designs aimed *to optimize sample size:* cross-over designs and enriched studies are methods used to minimize intersubject variability and maximize the expected differences between study groups, respectively, and thus reduce the number of patients needed to reach conclusions.

- Designs aimed *to shorten study duration* by the use of sequential designs, which analyze data as acquired and allows taking decisions as soon as evidence supports efficacy/futility, avoiding unnecessary experimental exposures.

- Designs aimed to *redefine and correct study designs* to maximize efficiency and minimize risks derived from deviations from wrong initial assumptions. Adaptive designs allow corrections to different aspects of the study design using acquired information generated during the study conduct.

- Designs aimed to *integrate results* throughout the use of Bayesian methods or modelling, allow incorporating previous information and making inferences taking profit of all the available knowledge.

There are no specially recommended methods for the design, carry out or the analysis of clinical trials in small populations. The designs outlined may represent different approaches to increase the efficiency of clinical trials. These methods can help in addressing many of the challenges posed by clinical developments of orphan drugs and are relevant to small clinical trials but also applicable to large studies.

The need for statistical efficiency should in all cases be weighed against the need for clinically relevant/interpretable results; the latter being the most important. However, relaxing statistical requirements should be left as a very last resource, when obtaining controlled evidence on the efficacy and safety of a new treatment is not possible by any other mean (52).

2.8. Development of alternative methods

Most statistical design and analysis methods for clinical trials, including those that might be more suitable for reduced samples, have been developed in the setting of confirmatory trials with relatively large sample sizes. Results obtained with these methods in wide populations may not be similar when applied to evaluate therapies in small populations. However, situations where limited number of patients could potentially be enrolled in the trial, raise specific statistical challenges and can lead to poorly designed studies and slower marketing approval of orphan drugs (53).

In the light of these challenges, in 2013 the EC sat up a unique call for new methodologies for clinical trials for small population groups within the FP7-HEALTH-2013-INNOVATION-1 framework (54). The objective of the research was to develop new or improved statistical methodologies for clinical trials for the assessment of treatments for small population groups, in particular for rare diseases or personalized (stratified or individualized) medicine. Research was expected to be multidisciplinary and involve all relevant stakeholders including industry and patient advocacy groups as appropriate.

ASTERIX (Advances in Small Trials dEsign for Regulatory Innovation and eXcellence, FP7 HEALTH 2013 – 603160 (55)) is one of the 3 multinational projects funded by the EC with the aim to develop innovative approaches to adapt and assess clinical trials on small populations and rare diseases. The objectives of the ASTERIX Project were to develop design and analysis methods for single trials and series of trials in small populations, include patient-level information and perspectives in design and decision making throughout the clinical trial process, and to validate new methods and propose improvements for regulatory purposes.

Other 2 projects were IDEAL (Integrated Design and Analysis of small population group trials, FP7 HEALTH 2013 – 602552) (56) and InsPIRe (Innovative methodology for small population research) (57). The objective of the Ideal research was to produce methods of general applicability irrespective of indication by Integrated DEsign and AnaLysis of clinical trials in small population groups (IDeAl) through a multidisciplinary closely collaborating consortium of researchers from European universities, research institutes and industry. The Ideal consortium focused on assessment of randomization procedures, extrapolating dose-response information, investigation of adaptive designs, optimal designs in mixed models,

pharmacogenetics designs, simulation of clinical trials, genetic factors influencing the response, decision analysis and biomarker surrogate endpoints (58).
The Inspire Project was focused on Bayesian and decision-theoretic methods that formally enable comparison of the gain in information with the cost, both in economic and opportunity terms, of clinical experimentation, and assess how information from outside the trial can formally be incorporated in the design and decision-making processes (59).

The Universidad Autónoma de Barcelona collaborated as a partner in the ASTERIX Project, leading Working Package 5 (WP5), which was dedicated to the assessment of the applicability of the new methods developed within the project, and to provide a set of regulatory recommendations (55).

2.9. Justification of the project

The methodological approaches to study new treatments for rare diseases struggle with the need to conclude efficacy and safety and the difficulties of achieving statistical demonstration with conventional statistics in small populations (49). The EU legislation determines that market access to new treatments requires the same level of evidence for rare and highly prevalent diseases (Regulation (EC) 141/2000) (23). While patient's safety and best interests lead these provisions, these may delay the access to new or improved therapies for orphan rare disease populations and pose difficulties to developers that may discourage the research of new treatments.

A number of alternative methodologies have been developed in the last decades aimed to deal with the assessment of evidence in small populations (60). However, obstacles to their wide implementation into research in orphan or rare diseases have included the lack of predictability of regulatory requirements and the fact that alternative approaches may be regarded from the regulatory point of view as not compliant with standard evidence requirements (61). On the other hand, a number of reviews of the amount and quality of evidence that supported regulatory decisions on OD have been published in past years, and the potential risks of accelerated approval procedures based on limited data have been repeatedly highlighted (15,62,63).

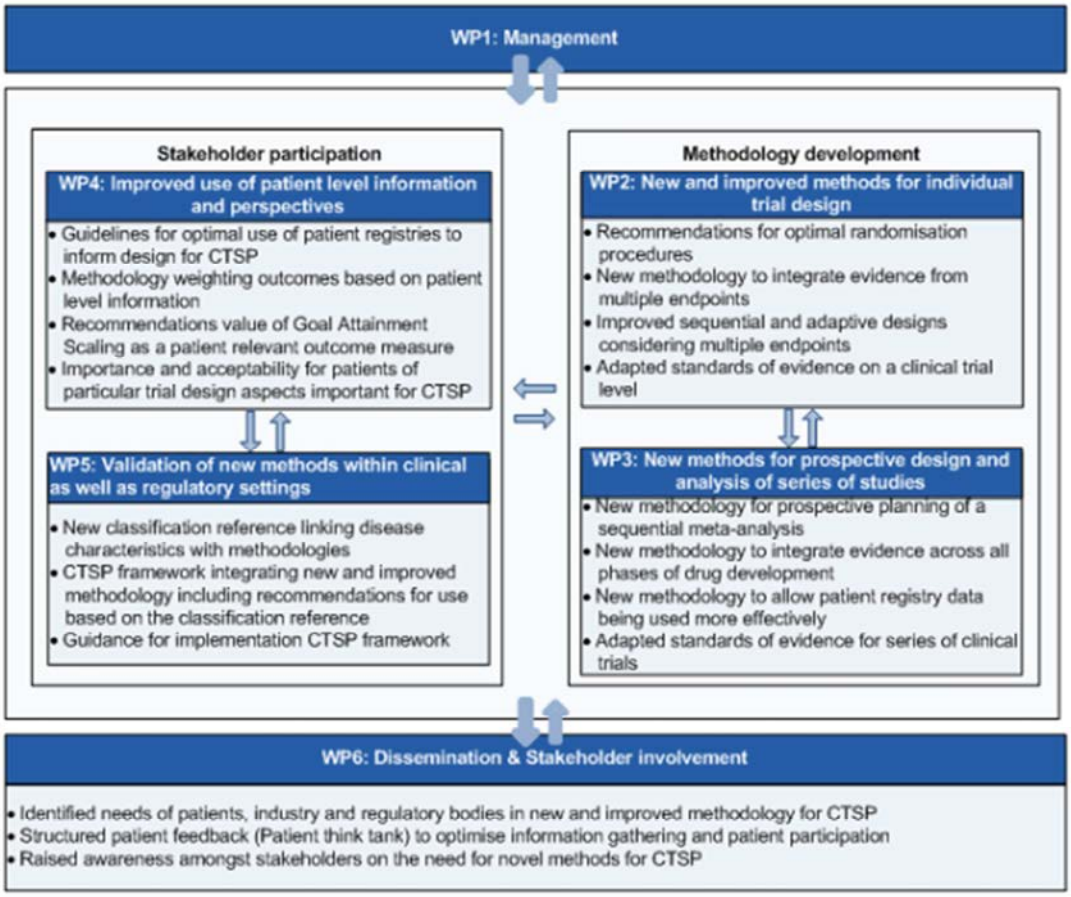
However, the applicability of the novel methodologies to the development programs in orphan conditions and the added value that these may have on the outcomes of regulatory evaluations when standard versus alternative approaches are applied to the same clinical situation have not been systematically evaluated. Applying novel methods to real examples may help to understand the risks and benefits of innovative approaches and may help to fill the acceptance gap between conventional and alternative methodologies, potentially normalizing their use in certain situations.

The present project was part of a collaborative European project funded by the EC (FP7-HEALTH-2013-INNOVATION-1 Grant-Agreement No. 603160. "Advances in Small Trials dEsign for Regulatory Innovation and eXcellence (ASTERIX)) (55) aimed to develop new methodologies for clinical trials for small population groups to increase efficiency of development programs. The focus was placed on ensuring patient safety while speeding patient access to new therapies.

The ASTERIX project

Our work in ASTERIX Project was integrated within that of six highly interactive and interdependent Work Packages (WPs) (64) (Figure 1).

Figure 1. ASTERIX Project: Structure and Relationship between Working Packages



Our Working Package (WP 5) had as a general objective to translate into clinical and regulatory recommendations the findings and results obtained by other Working Packages (WP2, WP3, and WP4). WP2 and WP3 were focused on statistical development of alternative approaches to the design or analysis of clinical trials in diseases characterized by difficulties to recruit enough subjects, and worked in collaboration with WP4, which was focused on integration of patient perspective into research. The project succeeded in obtaining a deepened insight in how methodology impacts assessment of evidence (65).

The achievements included:

- 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.

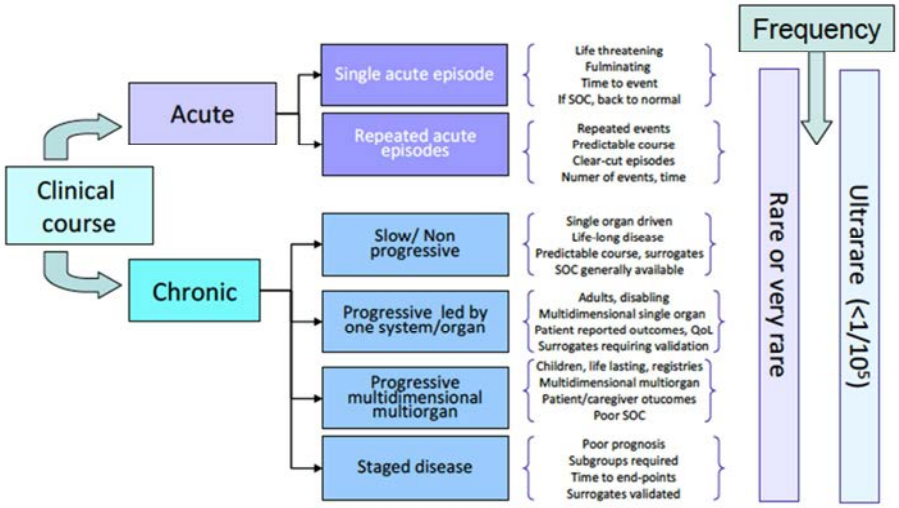
- Utter understanding and development of evidentiary standards for individual trials and drug development strategies, including the importance of randomisation and alternative strategies for exceptional circumstances.

A main challenge to reach the objective of translating these findings into clinical and regulatory recommendations was the lack of feasibility of issuing disease driven guidance, because of the huge number of existing orphan conditions. This problem was approached by proposing a clustering of rare conditions to structure the outcomes of other WPs of the ASTERIX project (42). This clustering aims to set a reference framework in order to offer specific methodology guidance, taking into account the clinical and methodological characteristics of any given orphan medical condition. In the proposed clustering the term *condition* is analogous to that of therapeutic indication and is not necessarily synonymous to disease; some diseases may have many conditions requiring different therapeutic approaches or interventions, which may require different methodological approaches to their study. The underlying principle is that the condition under investigation (i.e. the therapeutic indication sought for a given orphan disease), rather than the disease per se, is the key factor that determines the applicability of the different methodological approaches.

The aim of creating this clustering of medical conditions was to set a bridge between the current situation, where only the adopted general guidance for trials in small populations (49) is available and, in the other extreme, the unrealistic scenario attempting to provide a specified guidance for each separate orphan disease. Thus, since the goal of the clustering was to guide the design of clinical development of new treatments, the clustering is referred to conditions, and not to diseases.

The methodology and process for composition of the clusters has been previously described by our group (42,65) and is one of the main deliverables of WP5. Six disease clusters have been proposed (Figure 2).

Figure 2. Clustering of orphan and/or rare conditions based on clinical characteristics that determine the applicability of different research



The present work was focused on validating the applicability and value of the newly developed methods and alternative analyses within ASTERIX Project, and to integrate statistical methods into regulatory recommendations for the design or analysis of clinical trials in medical conditions with low prevalence, as clustered based on characteristics that allow shared methodological approaches for regulatory and clinical research purposes.

By doing so, we expected to improve guidance and ease the use in clinical research of newly developed methodologies, in order to reduce the uncertainty of clinical data supporting regulatory decision making and clinical practice in the field of rare diseases.

3. HYPOTHESIS AND OBJECTIVES

3.1. Hypothesis

Alternative methodological approaches to the study of rare or OD can be applied to the clinical development of new treatments, which may shorten the period of drug development and reduce the patient needs while keeping integrity and robustness of results.

Guidance to the application of such approaches to clusters of diseases or medical conditions sharing clinical characteristics determining the applicability of critical study design aspects may allow more specific regulatory guidance to researchers aiming to obtain a new marketing authorisation for medicinal products than those currently available.

3.2. Objectives

3.2.1. General Objectives

The main objectives of this work are:

- to validate the applicability of the alternative approaches developed within the ASTERIX Project to the design and analysis of clinical trials,

and

- to produce a set of recommendations on the applicability of new methodological approaches to the design and analysis of rare or OD based on the results obtained in the European project “Advances in Small Trials dEsign for Regulatory Innovation and eXcellence (ASTERIX)”, aimed to improve clinical trial guidelines both for regulatory purposes as well as for clinical researchers, by applying a clustering of medical conditions based on characteristics that allow shared methodological approaches.

3.2.2. Specific Objectives

1. Revision of available EU regulatory guidelines relevant for OD.

2. Selection of examples from each cluster of conditions to be used for the applicability and simulation exercises, with an assessment of its representativeness within the cluster.

3. Evaluation of the applicability of the methods developed within ASTERIX Project to the 6 clusters of medical conditions.

4. Simulation of clinical developments applying novel methods to real examples and analysis of the regulatory impact with regards to potential advantages or disadvantages in terms of study duration, simple size, and ethical aspects.

5. A proposal for a final set of basic recommendations to improve current EU guideline on clinical trials in small populations.

4. METHODS

The methods presented in this section have been previously described and submitted to the EC as part of the follow up and final deliverables of WP5 of the ASTERIX Project (65).

The work undertaken consisted of the evaluation of the applicability of the different methodologies and approaches developed by other Working Packages within the ASTERIX Project, to the six clusters of conditions developed by WP5. The applicability exercise was conducted in real life applications for marketing authorisation, based on data published in the EPARs. Other relevant methods identified outside of the ASTERIX project could be considered and tested, if deemed appropriate.

Subsequently, simulations of alternative development plans applying the novel methods were conducted in selected cases within each cluster of conditions and the added value was assessed. Based on the outcome of these initial analyses, and after establishing the current availability of EU regulatory guidance for the development of drugs in OD, an attempt to provide a final set of recommendations as suggestion for implementation in guidelines was done.

The task was structured following these steps:

1. Revision of the available EU regulatory guidelines relevant for OD.

2. Selection of the sample of EPARs for each of the six condition clusters.

3. Evaluation of the applicability of the methods developed by other WPs.

4. Assessment of the regulatory impact of the applicability of new methods to real examples (simulations).

5. Final set of basic recommendations as a suggestion for implementation in guidelines.

4.1. Revision of EU regulatory guidelines relevant for orphan diseases

Some of the EU guidelines currently available relate to the status of rare diseases and OMP like, for example, the Guideline on clinical trials in small populations (49). However, many gaps in this research field are still responsible for a high number of questions addressed to EMA (50). Seek of scientific advice for topics like the limitations imposed by the rarity of the disease and the collection of comprehensive data on safety and efficacy, or the acceptability of alternative designs remains high in general and in particular when a potential OMP is under investigation. This would underline the need for further regulatory guidance.

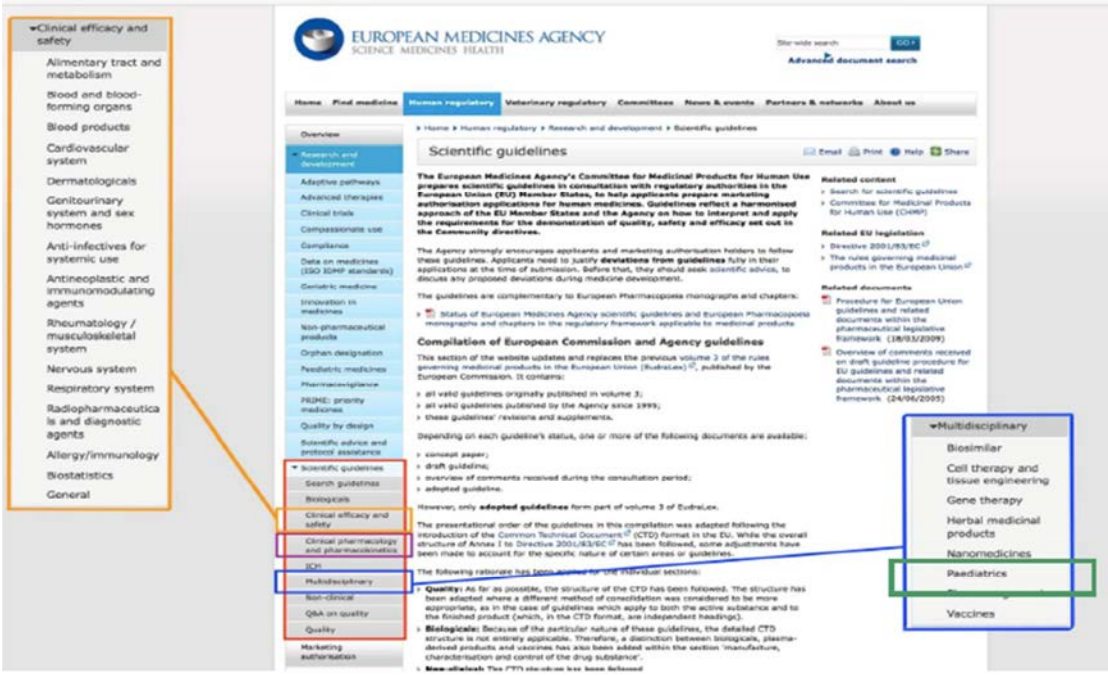
In order to accurately reflect the current situation, as a starting point in the exercise, the actual extent and availability of EMA guidance on the topic of OMP was analyzed. Thus, a systematic search of the available European regulatory guidance documents relevant to the development of new medicinal products for rare diseases was conducted to settle the current regulatory standards (48). A description and summary of the specific type of recommendations was done with the aim to identify any gaps or needs. This systematic review was deemed to enhance the effort of the WP5 within the framework of the ASTERIX Project and aimed to provide the basis

when proposing to the regulatory authorities an alternative approach with more specific and structured methodological recommendations for drugs targeting rare diseases that are under investigation.

The systematic search for guidelines related to rare diseases and the clinical development of OMP investigation was performed by searching the EMA website between April 2016-August 2017 (48).

Three subcategories of scientific guidelines were checked: the one referring to clinical efficacy and safety, the one referring to clinical pharmacology and pharmacokinetics (PK), and the clinical efficacy and safety documents located in the pediatrics category of the multidisciplinary guidelines (Figure 3) (65).

Figure 3. Search in EMA website for scientific guidelines



The *selection of the documents* was conducted as follows:

- A first identification of documents was based on the title, and in case of doubt the introductory part or the executive summary of each document was also consulted.

- Guidelines/Notes for guidance, addendum, questions and answers, points to consider, position papers and reflection papers were all included.

- Guidelines with title of warnings and core summaries of product characteristics and concept papers were not considered.

- Other documents were also revised, and their inclusion was decided case by case.
- The latest version that was in force was chosen. Efforts were also made to sift out the documents that were no longer into effect or had already been included in latest revision

of other documents. In case a guideline or a new version was in a draft stage, the research was focused on the draft, since often only minor modifications are done to the draft before the new version is adopted.

After the selection, *a review of the chosen guidelines* was performed. The procedure included the following steps:

- Documents in which a relation to the OMP investigation and to rare diseases was identified were subsequently categorised according to its general or specific applicability to OMP investigation.

- Information related to the conduct of clinical trials, specifically for the development of OMP, was thoroughly sought. In particular, clear reference to rare diseases, clear/specific OMP recommendations on the type of studies (e.g. efficacy/safety, exploratory/confirmatory/PK), on the trial design (e.g. randomised, controlled, duration), endpoints (e.g. primary, secondary), selection of subjects (e.g. inclusion/exclusion criteria, number), as well as on statistical considerations (e.g. plan for analysis, covariates, interim analyses, sample size reassessment), safety(e.g. adverse events), and paediatric population (e.g. special plans for the study in children) considerations were sought out. Recommendations concerning the need for post-marketing information were also searched.

Details about the results of the systematic search concerning OMP information extracted after the thorough review of a total of 71 documents were synthesized in a table containing the following information for each guidance document displayed in 12 columns: order number, EMA guideline reference number, applicability to OMP investigation (general, specific, or mixed), reference to rare diseases, presence of information about the type of studies, the study design, endpoints, patients selection, statistical considerations, safety considerations, pediatric considerations, and post-marketing data.

4.2. Selection of the sample of EPARs for each of the six conditions clusters

Given the large number of rare diseases described, in the range of 6,000 to 8,000 different diseases (6), issuing methodological guidance and validating methods for the study of every disease may become challenging. A clustering of rare medical conditions in 6 groups was previously proposed by our group, which was aimed to allow drawing specific methodological and regulatory recommendations applicable to types or groups of medical conditions, rather than to single disease models. The proposed clustering of medical conditions also served as a framework for testing the applicability and added value of new methods. In order to check the suitability of the newly developed statistical methods, and also the patient´s perspectives on study designs and conduction, relevant representative examples of medical conditions included in these six clusters were selected.

Overall approach and working plan:

All European Public Assessment Reports (EPARs) issued by the EMA for OMP with positive opinions since inception of the Orphan Regulation(23) and until December 2014 were identified from the EMA website.

Medical conditions were identified from all the designated orphan conditions as included in the authorised indications in the EPAR of each OMP.

Medical conditions were classified into the 6 ASTERIX clusters in replicate by 2 investigators. The frequency of the medical conditions was also classified as rare (≤ 5/10,000 and >1/100,000) and ultrarare (UR) (≤1/100,000).

The results of the two investigators were summarised and compared. Discrepancies were discussed until consensus was reached, emphasizing the search of conditions not fitting in at least one of the proposed clusters, or which could fit in more than one cluster (66).

A final listing of rare medical conditions was issued, including those for which two possible clusters could be assigned depending on the therapeutic approach.

The detailed planning of the work is summarised in Figure 4.

Figure 4. Working plan

Task
1. Identify from the European Medicines Agencies website all European Public Assessment Reports issued by the European Medicines Agencies for Orphan Medicinal Products, since inception of the Orphan Act and until December 2014.
2. Extract medical conditions from all the Designated Orphan Conditions as included in the claimed indications in the European Public Assessment Report of each Orphan Medicinal Product.
3. Classify in replicate by 2 investigators all medical conditions into the clusters and for the frequency of the medical conditions (rare vs ultrarare) according to the prevalence indicated by Designated Orphan Conditions.
4. Summarise and compare the classified conditions. Discussion of discrepancies until consensus, emphasising the search of conditions not fitting in at least one of the proposed clusters, or which could fit more than one cluster definition.
5. Issue a final listing of rare medical conditions, including those for which two possible clusters could be assigned depending on the therapeutic approach

A database had been created by WP 5 team as part of the ASTERIX project containing all available opinions on OMP issued by the EMA since Orphan Regulation No 141/2000 came into force up to December 2014 (67). This database was also used in this exercise.

In total 125 EPARs were evaluated for 85 rare conditions. The claimed indications in the EPAR of each OMP were extracted and the medical condition related to the indication was identified for classification.

The 85 medical conditions were classified independently by 2 different investigators, with no fixed pairs of investigators working on the same conditions, so that one criterion or pair of criteria did not systematically prevail over others, ensuring a high degree of consensus and reproducibility.

The data was then reconciled, and discrepancies discussed with the intervention of a third investigator until reaching a final consensus result.

The focus was on development plans and when deemed necessary, other sources were sought out. Negative opinions/withdrawals were not included given the very scarce amount of information publicly available.

The aim of this task was to select examples of EPARs representative of the conditions within each of the six clusters: acute single episodes (Cluster A), conditions with acute repeated episodes (Cluster B), chronic conditions with stable or slow progression (Cluster C), chronic progressive conditions lead by one system/organ (Cluster D), chronic progressive conditions led by multiple systems/organs (Cluster E), and chronic staged conditions (Cluster F). The selected EPARs were equally represented in spite of the fact that some scenarios were very frequent, e.g. 38/125 (30%) of all available EPARs in the EU by December 2014 correspond to staged conditions, and others very infrequent, e.g. 9/125 (7%) of all available EPARs concerned conditions with acute recurrent episodes. The selection was aimed to capture most of all possible scenarios, and not just being representative of the most frequent situations in the development of OMP. The following criteria for selection were used:

- Four EPARs for each of the six condition clusters. This number was considered sufficient to capture the diversity within the cluster, but it was acknowledged that exceptions were still possible.

- EPARs were selected regardless of type of authorisation, i.e. whether the approval was a regular approval, a conditional approval or an approval under exceptional circumstances.

- Since public detailed information on the development program and clinical trials for OMP is available for EMA positive opinions only, OMP with negative opinions and/or withdrawals from the evaluation procedure were not considered.

- Since extreme rarity of a given medical condition raises additional limitations to the recruitment potential, for each condition cluster at least one EPAR describing an ultra-rare condition (defined as affecting ≤1/100,000 persons in the EU) was selected, whenever possible.

- Only one EPAR per medical condition was included. The same drug could had been included more than once if developed for more than one indication, although none actually was.

- At least one repurposed drug per cluster was selected when possible, defined as a drug that was already in use for a different condition and for which a new authorisation was applied and granted for an orphan indication. Repurposed drugs may have different development approaches because part of the already available information may be extrapolated from former use to the new application.

From this set the focus was on the selection of the most recent applications, using the criteria described above, as this would have given the most up-to-date view on the methodologies used, aiming to select the majority of EPARs submitted after the guideline on clinical trials in small populations came into force (84/125 EPARs)(49).

When deemed necessary, specific or additional EPARs were revised for the evaluation of some specific methodology and this was decided on a case by case basis.

If information in EPARs was insufficiently detailed, FDA summary basis of approval, published original articles, and public clinical trial registries were consulted in order to have the necessary information for assessment (68–70).

4.3. Evaluation of the applicability of the methods developed by other WPs set

All novel methods that were developed within the ASTERIX project or tailored to small populations, and that had been reported in a published or (nearly) submitted manuscript by September 1st 2017, were included (20,71–83).

Manuscripts that discussed already existing methods or described a new perspective on an already existing method were excluded.

The methods were categorised in four main groups:

1. Six ‘innovative trial designs’, including: considerations on the use of delayed-start randomization (81), a method for interim analysis and stopping rules in multi-arm parallel trials (79), two methods for sample-size reassessment (one for adaptive survival trials (72), and a second one with a Bayesian approach for continuous end-points (78)), a method to optimize boundaries in group-sequential designs (73), and a method to weight prior information in Bayesian trials based on similarity of previous data (74).

2. One ‘level of evidence’ method, consisting of a set of recommendations to check if prior information can be used for inference allowing to relax the significance level in confirmatory trials, reducing sample size while controlling for certainty (77).

3. Four ‘study endpoints and statistical analysis’ methods, including: three methods to analyze multiple end-points (one for analysis of repeated measurements of multiple end-points (simultaneous inferences GEE models, was not available at the time of the exercise), one allowing conclusions for multiple co-primary endpoints even when not all meet statistical significance (71), and an exact non-parametric method for multiple

binary end-points (76)), and a new patient-guided measurement instrument aimed to standardize individual patient expectations on the treatment outcomes in conditions with heterogeneous clinical expression (75).

4. Two 'meta-analysis' methods, aimed both improving the management of heterogeneity estimators in meta-analysis of sparse-event studies (80,83).

Within each of these categories a variety of aspects of the methods were developed.

The requirements for use of each of the methods were identified based on the descriptions provided by WP 2, 3 and 4 and discussed with researchers of these teams whenever necessary. Based on these requirements, the conditions selected from EPARs were analysed for applicability of each of the method.

4.3.1. Method of evaluation of applicability and added value of novel methodology

Key characteristics on the studies that were used as pivotal evidence to support approval of orphan products were extracted from the EPARs and systematized through a list of items summarizing the key condition and treatment characteristics that might influence study design (Annex 1). The key features of the novel ASTERIX methods were summarized, including their prerequisites for applicability and potential advantages and disadvantages of the method.

For each studied method, once the data was extracted, interpreted and conclusions were drafted, these were sent for validation to the lead authors of the manuscripts describing the novel methods. Any disagreements between primary evaluators and authors were debated until consensus was reached.

In parallel, the list of items to extract from EPARs was completed, data were extracted for pivotal trials, including a summary of the condition, the trial characteristics needed to judge whether the pre-requisites for applying the novel methods could have been fulfilled, and any applicant's justification for choice of design elements and strategy, if available.

The prerequisites for applicability of the methods were assessed against the design characteristics of the pivotal studies and characteristics of the orphan conditions, and the design of the pivotal studies included in the EPAR. In addition, it was evaluated whether the method (if applicable) would have added value compared to the currently used method.

A pilot-testing was conducting on four studies reported in two EPARs (for OMP Savene® and Cayston®), and then the list of items with study and applicability details was refined (Annex 2).

If the method was considered applicable then, in addition, it was evaluated whether the method would add value as compared to the currently used method. The currently used method was used as a reference – rather than using one common standard as comparator, since it would be difficult to have a single gold-standard given the plethora of challenges associated with each condition and patient population. Also, the comparison to actual development best reflects the improvements that could be achieved for each scenario.

A two-step approach was used to determine whether or not the methods could have been applicable, and could potentially have added value to the overall development:

Step 1. *Static step*: evaluation of direct applicability without any adjustments to the original characteristics of the pivotal studies. Applicability was mainly driven by the (methodological) pre-requisites of the methods and whether these were fulfilled for the trials;

Step 2. *Dynamic step*: evaluation of applicability allowing for adjustments to the original setting or design of the studies, without changing the original objective and context of the development plan. Changes were made upon checking therapeutic guidelines, regulatory guidelines or any published article of a study in the same condition, in order to justify the applicability and potential improvement of the drug development program (61). For example, a secondary outcome could have been promoted to a primary outcome, or primary and secondary end-points could have been defined as multiple co-primary end-points, if this was clinically and methodologically appropriate and sound from a regulatory point of view.

4.3.2. Analysis, interpretation and synthesis of the results

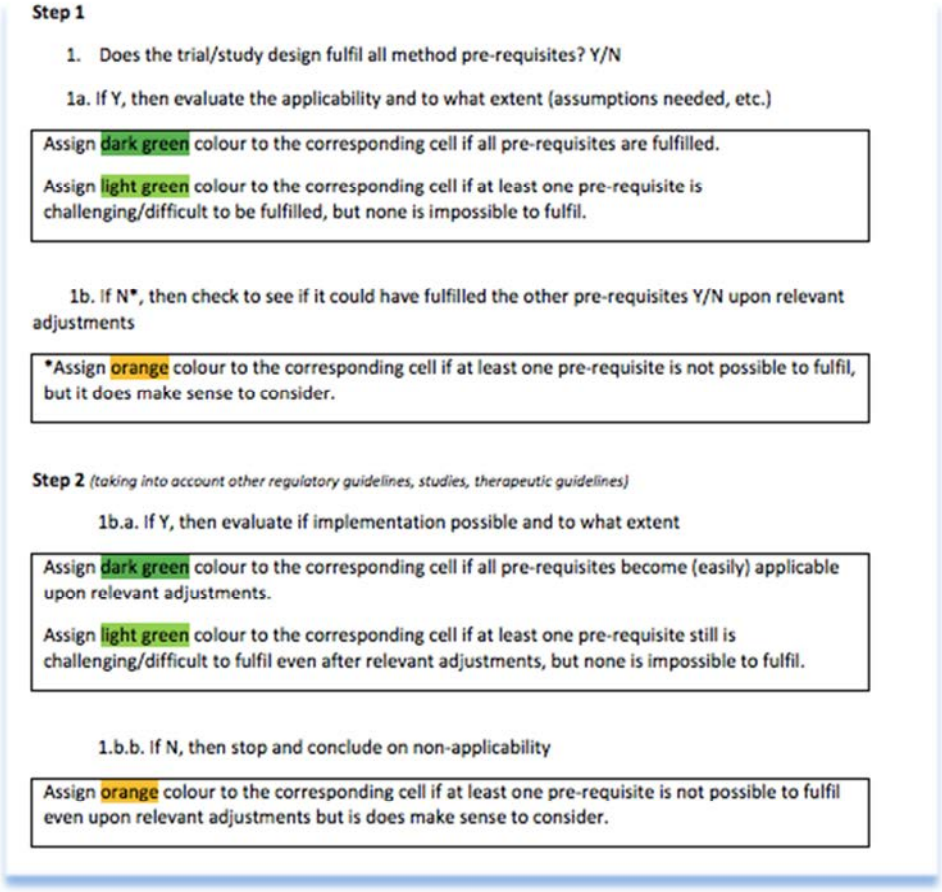
The methods pre-requisites and the characteristics of the pivotal trials were compared, and a decision tree structure was used for the evaluation of methods applicability (Figure 5) (61). The decision tree structure allowed measuring how applicable the methods were for each EPAR. The degree of applicability was categorized in four levels depending on fulfillment of pre-requisites, and a heat-map was used to summarise the results, where:

- ‘applicable’ was denoted by green colour;
- ‘may be applicable’ was denoted by light green colour;
- ‘no applicability’ was denoted by orange colour;
- ‘no possibility for application irrespective of changes’ was denoted by grey colour.

For step 1 if one of the pre-requisites was not fulfilled then non-applicability was concluded, while for step 2 if pre-requisites were not fulfilled, then relevant changes were explored before concluding on applicability or non-applicability.

The applicability was summarized numerically and visualized through the heatmap for the first (static) and second (dynamic) steps of the evaluation. Based on these heatmaps, recommendations on the use of the novel ASTERIX methodologies per cluster of conditions were drawn.

Figure 5. Decision tree structure for methods evaluation of applicability



4.4. Regulatory impact of applying new methods to real examples

Following the previous theoretical exercise, a more practical exercise was conducted. This consisted of designing alternative development plans for a number of selected real OMP examples following the recommendations on applicability and added value of the novel methodologies as concluded in the previous exercise. The aim of the alternative development plans was to address the main uncertainties identified by the CHMP of the EMA, as reflected in the EPARs, at the time of drawing a recommendation for the marketing authorisation of the due OMP. A systematic approach was followed for this exercise. For this purpose, a *drug development template document* with the following headings was created and completed for each example:

- Introduction: including background information on the disease, alternative treatment options, rationale for the development, as well as the actual scope of development.
- General investigation plan: objectives of development.
- Assessment of applicability of methods: representativeness of the example within the cluster, theoretical applicability of methods.
- Actual development plan for the chosen example.

- Alternative developments (scenarios 1-3).
- Analysis of the ethical, practical, and regulatory impact.
- Final recommendations.

4.4.1. Selection of EPARs and summary of applicability of methods

The initial plan within the ASTERIX project was testing the applicability and added value of new developed methods and approaches by using data from selected clinical trials in OD. We planned to obtain either publicly shared raw data from actual trials, or alternatively to approach companies for them to share pivotal databases for authorized products. However, access to actual raw data happened to be a long-lasting bureaucratic procedure that made it unfeasible granting access within the required tight timelines for the ASTERIX Project completeness. This was already anticipated at the time of project definition, and a contingency plan based on simulations was in place.

Thus, we decided to simulate as close as possible the real situation of planning a clinical development program, using the available information as summarized in the EPARs of OMP authorized by EMA. For this practical exercise, six out the 24 EPARs previously selected for the assessment of applicability of methods and representative of each of the medical condition clusters were chosen. The selection was based on the availability of recommendations for applicability of methods at the time the exercise was to be conducted, on the availability of enough information to identify and fully describe the key studies conducted 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 EPARs were scrutinized to find information concerning the main characteristics of the disease under investigation and on the existing treatment options, and these were duly summarized in the template document for each example. Further, the rationale for and the general objective of the actual development program were also critically revised, and these were reflected in the corresponding section of the template document.

Recommendations on applicability of methods for every of the 6 examples chosen, as determined in the previous exercise, were summarized in the template document with a description of requirements, potential advantages, and disadvantages.

The representativeness of the condition within the cluster was also qualitatively analyzed. This was deemed critical in order to guide the final recommendations for the cluster. The focus was placed on those aspects of the particular example under evaluation that might have an impact on the generalizability of conclusions reached with the given example, i.e. aspects that may differ from other conditions in the cluster and that could be determinant for the applicability of the methods. Examples of aspects that were considered are the course and speed of disease progression, the possibility to start development in adults, at least for the proof of concept studies, before moving to the paediatric population, whether the degree of rarity of the condition may pose difficulties for the uptake of sites, making adaptive methods applicable in spite of dealing with prevalent cases like in the case of chronic progressive clusters, the availability of

an effective standard of care (SOC) or treatment that may hugely differ between conditions within the same cluster, the existence of pharmacodynamics markers, or, for example, the existence of disease registries.

4.4.2. Analysis of actual development and regulatory uncertainties

A summary of the actual clinical development plan conducted was prepared for each of the conditions selected, based on the EPAR, publications of the pivotal trials, or information reported on the trials included in the EPARs at US and/or EU clinical trials registries (70,84), where required, and available.

Mimicking the process of clinical development planning (85), a simple target product profile was drafted based on the summary of products characteristics in order to focus the main goal of the development, and general objectives for the clinical development were then summarized. The actual clinical studies conducted within the clinical development program were listed, including summaries for those clinical studies that were relevant to support proof of concept, dose finding, pivotal confirmation or key supportive for MAA.

This description ended up with an analysis of the main regulatory uncertainties and weaknesses of the actual development as established in the benefit/risk section of the EPARs, so that the alternative methods could be oriented to address the areas of uncertainties and weaknesses through the simulations.

4.4.3. Simulation of alternative development programs

For each of the selected EPARs a revision of the disease identifying its clinical characteristics relevant to assess the therapeutic objectives in relation to the sought indication was conducted, which were then translated into research objectives. Considering the type of disease, the clinical relevance of the different possible endpoints was assessed, and the most efficient ones were selected to guide the clinical development. The number of relevant studies, their objectives, and the most efficient sequence for conducting the trials were evaluated. In addition, for each individual clinical trial the following aspects were systematically explored: study design, control group, number of study arms and type of assignment to treatment, sample size, endpoints, duration of recruitment period and individual patient follow-up, and statistical analysis. The advantages and disadvantages of the actual design and any potential variation by applying different approaches were evaluated based on the theoretical characteristics of the method, not conducting specific simulations.

At least one, but ideally up to 3, alternative clinical development plans were proposed for each of the conditions. This included a detailed study outline of the relevant trials where alternative methodologies were proposed. The alternative approaches were based on the evaluation of applicability of all methodologies described by WP2, 3 and 4, as described above, but also on other alternative methods described outside of the ASTERIX project, even if not novel, which were considered appropriate to improve the clinical development program of the drug for the condition and could represent a reasonable alternative.

4.4.4. Analysis of the regulatory impact and recommendations

Finally, the impact of applying novel methodologies was systematically assessed and described in a dedicated table (Table 1) for each of the alternatives proposed that were included in the template document. The impact on the following aspects was evaluated:

- *Practical impact*: e.g. on sample size requirements, time to study completion, ease recruitment, etc.
- *Methodological impact:* e.g. protection against type I and/or II errors, robustness of the evidence, internal validity, etc.
- *Ethical impact:* e.g. minimization of risks, maximizing access to an effective treatment, minimizing exposure to an ineffective treatment, considerations for patients input.
- *Regulatory impact:* e.g. availability of suitable information for benefit/risk assessment, external validity, informative on clinical relevance, credibility, etc. impact of each of the alternatives proposed was systematized in tables like the one that follows.

Table 1. Template for the analysis of the practical, regulatory, ethical, and methodological impact of applying an alternative method

Method assessed: Option 1: xxx	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements		
• May shorten time to study completion		
• May ease recruitment		
Statistical assessment:		
• Improves internal validity		
• Increases stability of estimates		
• Increases sensibility to changes		
• Compliant with predetermination		
• Consistency (discuss)		
• Robustness of method (discuss)		
• Protection against type I and II errors (discuss)		
Regulatory assessment:		
• Risk of bias and credibility		
• External validity (discuss)		
• Therapeutic positioning and comparisons		
• Informative on relevance and clinical impact		
• Enough information on safety		
• Suitable information for risk-benefit balance		
Ethical assessment:		
• May minimise risks		
• May maximise access to treatment		
• May minimise unnecessary exposure to ineffective treatments or placebo		
• Considers patient input		

Following the assessment of the practical, ethical, methodological and regulatory impact of the alternative development plans simulated for each example, a summary of recommendations on the most suitable novel methodologies with potential added value in the planning and conduct of a development plan for the very same condition under analysis were drawn and included in the template document.

4.5. Final set of basic recommendations as suggestion for implementation in guidelines

Finally, the internal validity and the generalizability of the results were discussed within the research group, including representation of experts in drug regulation. The focus was set in trying to determine to what extent the results could be translated into recommendations for the use of novel methodologies to the design, analysis, and interpretation of the clinical development results. Subject to the achievement of substantial agreement, a set of basic recommendations applicable to every cluster of conditions based on the simulations performed were to be set.

5. RESULTS

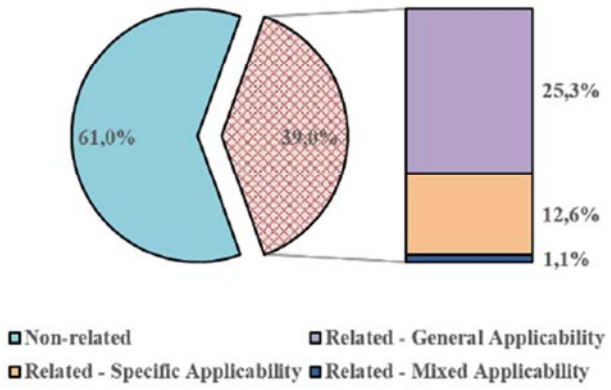
The results herewith presented have been previously summarized in the deliverables of the WP5 of ASTERIX Project and submitted to the EC as part of the periodic follow-up and final mandatory reports(65). *Additionally, some of these results have been published or accepted for publication in relevant medical journals (e.g. the Orphanet Journal of rare diseases* (42,61,86).

5.1. Revision of the EU regulatory guidelines relevant for orphan diseases

The systematic search conducted for EU guidelines related to rare diseases and development of the clinical trials part of an OMP investigation (86) raised the following results:

A total of 182 documents in the three subcategories explored of the EMA website fulfilled our search criteria. Among the whole spectrum of EMA's guidance documents, 71 (39%) were found to be applicable to the clinical development of OMP as it is shown in Figure 6 below. Documents found within two different subcategories were considered only in the one reviewed first to avoid duplications.

Figure 6. EMA documents related to OMP and type of applicability



The majority of documents (46 out of 71) were under the name "Guideline" or "Note for guidance" (Table 2).

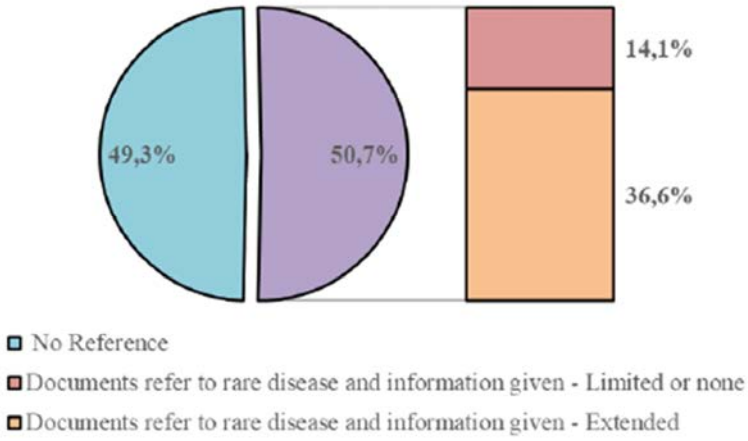
Table 2. Number and type of documents

Type of document	N°	N° of Drafts	N° of specific
Guidelines or Notes for guidance	46	6	15
Points to consider	4	1	1
Reflection Papers	5	1	1
Questions and Answers	5	-	-
Position Papers	1	-	-
Appendixes	3	-	1
Addendums	6	-	6
Others (i.e. collaborative approach)	1	-	1

Regarding the applicability of the guidance documents to OMP investigation, the systematic review revealed that 46 out of 182 (25.3%) documents were identified to have general applicability, 23/182 (12.6%) were lying in the category of specific applicability to OMP; 2/182 (1.1%) had both general applicability as well as specificity were revealed since reference to some subgroups of rare conditions was described in the same document of an otherwise not so rare condition (Figure 7).

Specifically, 36 of the 71 documents (50.7%) have references to rare diseases and/or OMP, whereas in 10 (14.1%) among those 36 the reference to rare diseases is accompanied with limited or none specific information on the characteristics which are considered important when designing and conducting a clinical trial in the orphan situation (Figure 7).

Figure 7. Reference to rare diseases in the documents related to OMP



Reference to paediatric information related to rare diseases was found in 21 of the 71 documents, and the need for post-marketing data for efficacy and/or safety in the OMP investigation was discussed in 17 out of 71 documents.

Detailed information referred to OMP in the 71 documents is presented in Annex 3.

5.2. Selection of the sample EPARs for each of the six condition clusters

None of the 26 OMP chosen had published data in any of different datasets that could have been consulted for details in case of insufficient information in the EPARs (68–70,84), so the only source of information were the EPARs.

Selection criteria described in Section 4.2. (Methods) were applied, and the following considerations were made:

- *Selection of EPARs with UR conditions*: In the set of 125 EPARs, there was only one ultra-rare disease in each of the following clusters of conditions: Cluster A: acute single episodes, Cluster B: acute recurrent episodes, Cluster C: chronic stable/slow progression and Cluster D: chronic progressive led by one system/organ. Thus, these

examples were directly selected in order to account for the ultra-rare diseases representation. There was no EPAR entering in the ultra-rare status in Cluster F: chronic staged condition, and therefore none could be selected. For the Cluster E: chronic progressive led by multiple systems/organs, 9 out of the 23 conditions were ultra-rare. Therefore, 2 of these ultra-rare EPARs were selected to adequately represent the ratio rare: ultra-rare in this category.

- *Avoid duplicity in drug products*: Moreover, in cases where the same disease could be assigned to more than one cluster of conditions and there were medicinal products targeting different aspects of the disease, and thus both EPARs could have been selected according to the study methods, only one was finally chosen in order to increase diversity. For example, cystic fibrosis was categorised in the main Cluster B: acute repeated episodes when targeting lung infection, and in Cluster D: chronic progressive led by multiple systems/organs when targeting other symptoms or modifying channel function. According to the study methods, EPARs for different drugs targeting different aspects of the disease and thus belonging to different clusters could both have been selected. However, in this exercise only the EPAR of *Kalydeco®*, a modifier of the channel function affected in cystic fibrosis, was the one finally selected.

- *Selection of repurposed drugs*: In addition, the inclusion of authorized drugs repurposed for new indications was intended, as special cases for which evidence previously authorized or data on the use of the drug in a larger population for a non-orphan indication may be extrapolated to the new condition and contribute to the evaluation of the safety profile of the medicinal product. However, there were no options for repurposed drugs in the cluster of chronic progressive conditions affecting multiple systems/organs.

- *Choice of recent dossiers*: The aim was to select the most recent applications, and EPARs submitted after the guideline on clinical trials in small populations came into force, but this was not always possible. In Cluster B: acute repeated episodes, in 3 out of 4 selected EPARs the applications had been submitted before the date when the guideline on trials in small populations had come into effect.

- *Exclusion of diagnostic drugs*: Medicinal products included happened to be used for the treatment and not for diagnosis or prevention of the disease. Overall only few exceptions to the rule occurred, e.g. Gliolan® was used for intra-operative photodynamic diagnosis of residual glioma. These were not selected based on the criteria outlined above, and no exceptions were made in this case.

In addition, some exceptions had to be implemented in this process:

- *Diverse therapeutic areas*: All EPARs initially selected in Cluster F: chronic staged conditions were neoplasms. In order to increase diversity, 2 of these examples were replaced by recent staged conditions from a different therapeutic field such as Opsumit® to treat pulmonary hypertension and *Revlimid®* for the treatment of myelodysplastic syndromes.

Finally, some misclassifications of the set of EPARs occurred, that did not consider the strong emphasis done on the pair "disease plus therapeutic indication" when referring to medical condition. Thus, some of the clusters assigned to the EPARs had to be amended, so that 2 EPARs were reclassified from Cluster E: chronic progressive led by multiple systems/organs to Cluster B: acute recurrent episodes. Specifically, the changes include the following:

- Sicklos® was firstly classified in Cluster D: chronic progressive condition led by one system/organ, considering the targeted disease in isolation, but according to the EPAR it had been developed for the prevention of recurrent painful vaso-occlusive crises; so it was moved to Cluster B: acute recurrent episodes.

- Orphacol® was firstly classified in Cluster E: chronic progressive led by multiple systems/organs for the very same reasons but was subsequently reclassified in Cluster D: chronic progressive led by one system/organ since all the assessments and measurements referred to liver functionality and damage.

- Tracleer® was seeking an indication for prevention of new digital ulcerations, but its effect on other organ/systems commonly affected in Systemic Sclerosis were not investigated; so, it was reclassified from Cluster E:chronic progressive led by multiple systems/organs to Cluster B: acute recurrent episodes.

- Kalydeco® and Vyndaqel® were selected to replace the two examples removed from Cluster E: chronic progressive led by multiple systems/organs. The first one is used in order to improve the defective opening of the chloride channels in cystic fibrosis that can affect different functions (e.g. lungs, nutrition). The second one is indicated for familial amyloid polyneuropathy, where different functions and organs are affected.

Therefore, two additional EPARs were added to Cluster E: chronic progressive led by multiple systems/organs to ensure that there were at least 4 EPARs per cluster, leading finally to 26 EPARs in total.

A final list of selected EPARs was generated, after ensuring that the focus was put on the actual indication for which the medicinal product was developed and a marketing authorisation was finally granted. This final list included 4 EPARs for each cluster, as initially provisioned, except for Cluster B, which had a total of 6 EPARs due to the fact that two of the reclassified EPARs had already been assessed. The final list of the selected EPARs is presented below in Table 3.

Table 3. Selection of EPARs for evaluation of methodologies

Condition	Drug brand name	EPAR*	Date opinion	R/U**	Repurposed
A. Acute: single episodes					
Antracycline extravasation	Savene®	30	2006	U	No
Patent ductus arteriosus	Pedea® *(ibuprofen is repurposed)*	17	2004	R	Yes
Hepatic venooclusive disease	Defitelio®	105	2013	R	No
Tuberculosis	Sirturo®	108	2014	R	No
B. Acute: recurrent episodes					
Cryopirine periodic syndromes	Ilaris®	70	2009	U	No
Gram negative lung infection in cystic fibrosis	Cayston® *(aztreonam is repurposed)*	67	2009	R	Yes
Narcolepsy	Xyrem®	22	2005	R	No
Dravet syndrome	Diacomit®	39	2001	R	No
Sickle cell disease	Sicklos®	46	2007	R	No
Systemic sclerosis – digital ulcerations	Tracleer®	5	2009	R	No
C. Chronic: stable/slow progression					
Short bowel syndrome	Revestive®	92	2012	R	No
Adrenal insufficiency	Plenadren® *(hydrocortisone is repurposed)*	84	2011	R	Yes
Thrombocytemia	Xagrid®	19	2004	R	No
Deficit of lipoprotein lipase	Glybera®	94	2012	U	No
D. Chronic: progressive, one system/organ					
Nocturnal Paroxysmal haemoglobinuria	Soliris®	44	2007	R	No
Wilson's disease	Wilzin®	18	2004	R	No
Congenital errors of bile synthesis	Orphacol® *(colic acid is repurposed)*	104	2013	U	Yes
Gastrointestinal stromal tumours	Glivec® *(many other indications)*	3	2009	U	Yes
E. Chronic: progressive, multiple systems/organs					
Fabry disease	Fabrazyme®	1	2001	U	No
Cystic fibrosis – modify channel function	Kalydeco®	90	2012	R	No
Familial amyloid polyneuropathy	Vyndaqel®	85	2011	R	No
Gaucher disease	Zavesca®	8	2002	R	No
F. Chronic: staged condition					
Renal carcinoma	Afinitor® *(everolimus is repurposed)*	66	2009	R	Yes
Pulmonary hypertension	Opsumit®	106	2013	R	No
Indolent non-Hodgkin lymphoma	Litak®	15	2004	R	No
Myelodysplastic syndrome	Revlimid®	43	2013	R	No

*Internal number of EPARs per condition category (in total 85 conditions, 125 EPARs)

**R: rare prevalence *rare ≤5/10,000 and >1/100,000*; U: ultra-rare prevalence ≤1/100,000

5.3. Evaluation of the methodology set applicability

5.3.1. Description of the methods evaluated

The following table (Table 4) summarizes the key features of the 14 novel ASTERIX methods, their prerequisites for use, and potential advantages and disadvantages from a theoretical point of view (61).

Table 4. Overview of the methods that were evaluated

Description of the method	Requirements for use of the method	Potential advantages	Potential disadvantages
INNOVATIVE TRIAL DESIGNS			
Multi-arm group sequential designs with a simultaneous stopping rule			
A design with 3 arms or more, with planned interim analyses with a simultaneous stopping rule. This rule aims to detect at least one efficacious treatment out of all tested arms, and stops the trial for all arms as soon as for at least one treatment arm (in-)efficacy is proven, when the critical boundaries of efficacy or inefficacy are crossed.	▪ At least 3 arms including control (placebo) ▪ At least 1 interim analysis ▪ Time to outcome faster than accrual/enrolment rate ▪ Developed for continuous endpoints, transportable to other types (i.e. binary)	▪ More robust evidence on efficacy or for dose selection and/or posology, ▪ Avoidance of a placebo arm in some instances ▪ Preservation of alpha and power ▪ Efficient use of available patients	▪ Not applicable to historically/externally controlled studies ▪ More complex trial conduct ▪ Interim could result in an overall longer trial ▪ The potential to marginally miss a second efficacious intervention
Sample size reassessment and hypothesis testing in adaptive survival trials			
In general, this design allows a sample size reassessment during a trial where the primary outcome is the occurrence or absence of an event. The sample size reassessment will be done during an interim analysis aiming to reach sufficient evidence to reject or accept the efficacy of the intervention. This paper discusses major drawbacks of a fully unmasked sample-size recalculation, i.e. a decision based on all available efficacy and safety data, are potential intentional changes in the behaviour of the investigators, and the potential impossibility to include all patients in the final analysis and propose a test statistic for inclusion of all patients.	▪ In case the sample size re-assessment is unmasked. ▪ Time to outcome faster than accrual rate ▪ At least one interim analysis	▪ Increased precision for sample size reassessment ▪ Preservation of type I error ▪ Inclusion of all (more patients) in the final test statistic	▪ Logistically resource-wise more demanding
Sequential design for small samples starting from a maximum sample size			
Using a group sequential design, an analysis will be performed before the trial is finished, based on the available data collected at that (pre-defined) moment. The aim of this design is to pick up large benefits, lack of benefit and safety signals earlier. The proposed method uses the maximum available sample size as a starting point for planning the study, taking into account the desired chance to pick up a therapeutic effect if it really exists, and then continues with the refined calculations of the limit boundaries. This method determines the optimal number of interim analyses to be performed, while keeping the chance low of concluding that a treatment works - while in real life it does not work.	▪ Needs to start from maximum sample size that can be recruited ▪ At least 1 interim analysis ▪ Time to outcome faster than accrual rate	▪ Increased precision for the prior for treatment effect size estimates, and thereby increased precision for the adjustment of boundaries ▪ increased safety surveillance ▪ Optimised use of available patient pool ▪ Efficacy results obtained earlier may lead to less approval procedural time	▪ More interim analyses will provide extra work ▪ Sufficient level of evidence, but not overwhelming

Table 4 (Cont.) Overview of the methods that were evaluated

Description of the method	Requirements for use of the method	Potential advantages	Potential disadvantages
INNOVATIVE TRIAL DESIGNS			
Bayesian sample size re-estimation (to appear)			
Bayesian statistics, use probability distributions, often including a probability of the belief in the intervention before the start of the trial (the prior). For normally distributed outcomes, an assumption for the variance needs to be made to inform the sample size needed, which is usually based on limited prior information, especially in small populations. When using a Bayesian approach, the aggregation of prior information on the variance with newly collected data is more formalized. The uncertainty surrounding prior estimates can be modelled with prior distributions. The authors adapt the previously suggested methodology to facilitate sample size re-estimation. In, addition, they suggest the employment of power priors in order for operational characteristics to be controlled.	▪ At least 1 interim analysis ▪ Randomisation ▪ 1 control and 1 experimental arm ▪ Developed for continuous endpoints, transportable to other types of outcomes	▪ Fewer patients needed ▪ Optimised use of accumulated knowledge from previous studies ▪ Control of type I error	▪ Extra patients needed in case of effect size overestimation
Dynamic borrowing using power priors that control type I error (to appear)			
In rare diseases, where available data is scarce and heterogeneity between trials is less well understood, the current methods of meta-analysis fall short. The concept of power priors can be useful, particularly for borrowing evidence from a single historical study. Such power priors are expressed as a parameter, which in most situations has a direct translation as a fraction of the sample size of the historical study that is included in the analysis of 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. Therefore in this paper a new, simple method of estimating the power parameter in the power prior formulation is suggested, suitable when only one historical dataset is available. This method is based on predictive distributions and parameterized in such a way that the type I error can be controlled, by calibrating the degree of similarity between the new and historical data.	▪ Essential to have robust data from ideally previous similar studies ▪ Developed for normal responses in a one or two group setting, but the generalization to other models is straightforward	▪ Fewer patients ▪ Optimised use of accumulated knowledge from previous studies ▪ Control of type I error	▪ Extra patients needed in case of effect size overestimation

Table 4 (Cont.) Overview of the methods that were evaluated

Description of the method	Requirements for use of the method	Potential advantages	Potential disadvantages
INNOVATIVE TRIAL DESIGNS			
Delayed-start randomisation (to appear)			
In a delayed start randomization design the moment of start of the intervention is randomized. Using blinding, the patients will not know at what moment the intervention will start for them, reducing the placebo effect and any potential effect of the patients knowing he is receiving an active intervention. The change from the baseline measurement within each patient will be used as an outcome measure.	▪ The comparator needs to be placebo ▪ Intervention needs to have lasting response	▪ All patients will receive treatment eventually ▪ The switch moment may provide extra information	▪ Delay in some patients receiving treatment, compared to a single arm trial
LEVEL OF EVIDENCE			
Evidence, eminence and extrapolation			
In small populations, a full independent drug development program to demonstrate efficacy may not be ethical, feasible or necessary. Extrapolations of evidence from a larger population to the smaller target population are widely used to support decisions in this situation. This paper discusses clinical trial designs which make use of prior knowledge on efficacy for inference. A framework based on prior beliefs is formulated to investigate whether the significance level for the test of the PEP in confirmatory trials can be relaxed, and the sample size reduced, while controlling a certain level of certainty about the effects. The authors show that point-priors have some favourable properties over other types of priors. The crucial aspect to be made clear is the prior belief in the possibility of extrapolation from a larger population to the target population.	Factors that influence the possibility for extrapolation: ▪ Same underlying mechanism of action, similarity of response to treatment, similar dose-response relationship so the mechanism is translatable to the target population? ▪ Same disease symptoms in adults and children, regarding similarity of disease progression. ▪ Timing of the paediatric trial compared to the adult trial should allow extrapolation ▪ Repurposed drug or not or extension of indication	▪ Optimised use of available evidence. ▪ Reduction of sample size	▪ Difficulty lies in its novelty and application.

Table 4 (Cont.) Overview of the methods that were evaluated

Description of the method	Requirements for use of the method	Potential advantages	Potential disadvantages
STUDY ENDPOINTS AND STATISTICAL ANALYSIS			
Simultaneous inference for multiple marginal GEE models (to appear)			
A framework is proposed for using generalized estimating equation models for each endpoint marginally considering dependencies within the same subject. The asymptotic joint normality of the stacked vector of marginal estimating equations is used to derive Wald-type simultaneous confidence intervals and hypothesis tests for linear contrasts of regression coefficients of the multiple marginal models. The small sample performance of this approach is improved by adapting the bias correction proposed by Mancl and DeRouen to the estimate of the joint covariance matrix of the regression coefficients from multiple models. As a further improvement a multivariate t-distribution with appropriate degrees of freedom is specified as reference distribution. Alternatively, a generalized score test based on the stacked whom correspondence should be addressed estimating equations is derived. By means of simulation studies, control of type I error rate for these methods is shown even with small sample sizes and also increased power compared to a Bonferroni multiplicity adjustment. The proposed methods are suitable to efficiently use the information from dependent observations of multiple endpoints in small-sample studies. If simultaneous confidence intervals for two or more endpoints are of interest, this approach can be used. Additionally, an R software package has been developed (`mmmgee') for computational implementation of this framework.	▪ Repeated measurements	▪ Robust evidence from longitudinal data.	▪ Technically more complex

Table 4 (Cont.) Overview of the methods that were evaluated

Description of the method	Requirements for use of the method	Potential advantages	Potential disadvantages
STUDY ENDPOINTS AND STATISTICAL ANALYSIS			
Fall-back tests for co-primary endpoints			
Usually, when the efficacy of an intervention is measured by co-primary endpoints, efficacy may be claimed only if for each endpoint an individual statistical test is significant. While this strategy controls the type I error, it is often very conservative, and does not allow for inference if only one of the co-primary endpoints shows significance. This paper describes the use of fall-back tests. They reject the null hypothesis in exactly the same way as the classical tests, with the advantage that they allow for inference in settings where only some of the co-primary endpoints show a significant effect. Similarly to the fall-back tests defined for hierarchical testing procedures, these fall-back tests for co-primary endpoints allow continuing testing, even in the primary objective of the trial was not met.	▪ At least 2 co-primary endpoints ▪ One test per endpoint	▪ No need for hierarchical pre-specification and testing of multiple co-primary endpoints. ▪ Improved statistical testing (more chances to detect one dimension of treatment effect and benefit).	▪ Potentially more patients needed
Optimal exact tests for multiple binary endpoints			
In confirmatory trials with small sample sizes, hypothesis tests that developed for large samples - based on asymptotic distributions - are often not valid. Exact non-parametric procedures are applied instead. However, exact non-parametric procedures are based on discrete test statistics and can become very conservative. If adjustment for multiple testing, they become even more conservative. Exact multiple testing procedures are proposed, for the setting where multiple binary endpoints are compared in two parallel groups. Based on the joint conditional distribution of the test statistics of Fisher's exact test, the optimal rejection regions for intersection hypothesis tests are constructed. To efficiently search the large space of possible rejection regions, an optimization algorithm is proposed based on constrained optimization and integer linear programming. Depending on the objective of the optimization, the optimal test yields maximal power under a specific alternative, maximal exhaustion of the nominal type I error rate, or the largest possible rejection region controlling the type I error rate. Applying the closed testing principle, the authors construct optimised multiple testing procedures with strong familywise error rate control. In addition, they propose a greedy algorithm for nearly optimal tests, which is computationally more efficient.	▪ Multiple dichotomous/binary outcomes ▪ Two up to five endpoints ▪ Gain is most in very small sample sizes (1 to 50 per group) ▪ A priori (optimal) rejection region (defined) ▪ Prior distribution of effect sizes	▪ Optimised multiple testing procedure ▪ Maximal power use of the statistical test ▪ Control of family-wise error rate (FWER) ▪ Robust evidence ▪ Excellent for small sample sizes	▪ Potentially more patients needed

Table 4 (Cont.) Overview of the methods that were evaluated

Description of the method	Requirements for use of the method	Potential advantages	Potential disadvantages
STUDY ENDPOINTS AND STATISTICAL ANALYSIS			
Goal attainment scaling (GAL)			
Goal Attainment Scaling is a measurement instrument that measures the attainment of different goals of patients in a standardized way. The goals are measured in the same way for every patient, but the content of the goals can be different between patients. To apply GAL, the caregiver and the patient sit together to decide what the goals of the patient are, and how they can be defined in five levels. Next, the patient receives the intervention (preferably blinded). Then after the intervention the patient and doctor assess how well the goals have been attained. The different content of the goals for different patients makes that GAL can be used in groups of patients who all have different complaints, which is often the case in rare diseases. Another advantage is that it is very sensitive to change.	▪ Essential that the PEP is not relevant for all patients ▪ Heterogeneous disease course with stable baseline ▪ It has to be actual treatment not prevention ▪ Randomisation ▪ Measurement relevant at functional level	▪ The goals are individually defined by patients and chosen per patient, hence customised measurement of therapeutic effect. ▪ Time-demanding aspect, needed for detailed construction and definition of goals, may be less of a concern when there are a (very) limited number of available patients.	▪ Time-consuming to set goals individually per patient ▪ Choice of goals must be realistic ▪ Hawthorne effect
META-ANALYSIS			
Prior distributions for variance parameters in sparse-event meta-analysis (to appear)			
The small sample sizes in rare diseases make it particularly valuable to pool the data of small studies in a meta-analysis. When the primary outcome is binary, small sample sizes increase the chance of observing zero events. The frequentist random-effects model is known to induce bias and to result in improper interval estimation of the overall treatment effect in a meta-analysis with zero events. Bayesian hierarchical modelling could be a promising alternative. Bayesian models are known for being sensitive to the choice of between-study variance (heterogeneity) prior distributions in sparse settings. In a rare disease setting, only limited data will be available to base our prior on, therefore, the need to identify priors with robust properties is crucial. This paper shows that the Uniform (-10; 10) heterogeneity prior on the log (T^2) scale shows appropriate 95% coverage and induces relatively acceptable under/over estimation of both the overall treatment effect and heterogeneity, across a wide range of heterogeneity levels. We illustrate the results with two examples of a meta-analysis with a few small trials.	▪ >=2 RCTs ▪ Same endpoint in at least two trials, from which one PEP ▪ Binary endpoint(s) ▪ Sparse events ▪ Not prerequisite but patients allocation ratio ideally 1:1 ▪ treatment effect size estimates reported in harmonized (or harmonisable) manner ▪ Not prerequisite but ideally equally allocated number of patients per study	▪ Optimised use and variance estimation in a sparse-event MA. ▪ Quicker and optimal selection of appropriate heterogeneity priors distributions.	

Table 4 (final) Overview of the methods that were evaluated

Description of the method	Requirements for use of the method	Potential advantages	Potential disadvantages
META-ANALYSIS			
Heterogeneity estimators in zero cells meta-analysis (to appear)			
When a meta-analysis consists of a few small trials that report zero events, accounting for heterogeneity in the estimation of the overall effect is challenging. In practice, the data poses restrictions on the meta-analysis method employed that lead to deviations from the pre-planned analysis, such as the presence of zero events in at least one study arm. Estimators that performed modestly robust when estimating the overall treatment effect across a range of heterogeneity assumptions were the Sidik-Jonkman, Hartung-Makambi and improved Paul-Mandel. The relative performance of estimators did not materially differ between making a predefined or data-driven choice. The simulations confirmed that heterogeneity cannot be estimated reliably in a few small trials that report zero events. Estimators whose performance depends strongly on the presence of heterogeneity should be avoided. The choice of estimator does not need to depend on whether or not zero cells are observed.	▪ >=2 RCTs ▪ Same endpoint in at least two trials, from which one PEP ▪ Binary endpoint(s) ▪ Sparse events ▪ Not prerequisite but patients allocation ratio ideally 1:1 ▪ treatment effect size estimates reported in harmonized (or harmonisable) manner ▪ Not prerequisite but ideally equally allocated number of patients per study	▪ Quicker and optimal selection of heterogeneity estimator in a sparse-event meta-analysis	▪ Niche method and does not cover all heterogeneity estimators

5.3.2. Potential applicability of methods based on information from actual trials

Applicability of methods to the 26 EPARs

The assessment of applicability showed the following results:

- In the first static step, considering just the actually conducted development plan, it was found that all individual methods were directly applicable to a minimum of 1 (4%) up to 9 (35%) of the 26 EPARs. Overall each method was applicable to a minimum of 1 (17%) and a maximum of 5 (83%) of the 6 clusters.

- In the second dynamic step, where some adjustments to the development plan could be considered, it was found that the individual methods were applicable in 1 (4%) up to 17 (65%) of the EPARs, and to a minimum of 1 (17%) and a maximum of 6 (100%) out of the 6 clusters (Table 5).

Table 5. Percentage of EPARs where the methods are applicable

METHOD	Applicability in percentage of EPARs	Applicability in percentage of EPARs
	Static step 1	Dynamic step 2
Long-short outcomes	7,7% [2/26]	38,5% [10/26]
Evidence, eminence and extrapolation	34,6% [9/26]	46,2% [12/26]
Heterogeneity estimators	3,85% [1/26]	3, 85% [1/26]
Prior distributions for variance parameters in sparse-event meta-analysis	3,85% [1/26]	3, 85% [1/26]
Delayed-start randomisation	11,5% [3/26]	11,5% [3/26]
Sample size reassessment and hypothesis testing in adaptive survival trials	34,6% [9/26]	57,7% [15/26]
Multi-arm group sequential designs with a simultaneous stopping rule	23,1% [6/26]	57,7% [15/26]
Sequential designs for small samples	30,8% [8/26]	65,4% [17/26]
Bayesian sample size re-estimation using power priors	11,5% [3/26]	50% [13/26]
Dynamic borrowing through empirical power priors that control type I error	15,4% [4/26]	50% [13/26]
Fallback tests for co-primary endpoints	15,4% [4/26]	50% [13/26]
Optimal exact tests for multiple binary endpoints	3,85% [1/26]	30,8%[8/26]
Simultaneous inference for multiple marginal GEE models	19,2% [5/26]	23,1%[6/26]
Goal Attainment Scaling	30,8% [8/26]	30,8% [8/26]

Summary of applicability of methods by cluster

The number of EPARs for which the new methods were tested and the results on the number of tests that deemed applicable in each step, overall and for UR and repurposed (RP) drugs, is summarized in Table 6.

Table 6. Summary of applicability by clusters

	Sample of EPARs tested			Applicability in Step 1			Applicability in Step 2		
	N (%) EPARs	N (%) EPARs UR	N EPARs (N tests) N UR/RP (N tests)	N yes	N (%) UR	N (%) RP	N yes	N (%) UR	N (%) RP
Acute single episodes	23 (18%)	1 (6%)	4 (52) 1 (13) UR&RP	2/52 (4%)	0/13 (0%)	0/13 (0%)	14/52 (27%)	3/13 (23%)	1/13 (8%)
Recurrent acute episodes	9 (7%)	2 (13%)	6 (78) 1 (13) UR&RP	18/78 (23%)	2/13 (15%)	6/13 (46%)	41/78 (53%)	9/13 (69%)	8/13 (62%)
Chronic slow or non-progressive condition	13 (10%)	1 (6%)	4 (52) 1 (13) UR&RP	7/52 (13%)	0/13 (0%)	3/13 (23%)	16/52 (31%)	0/13 (0%)	4/13 (31%)
Progressive led by one organ-system	19 (15%)	3 (19%)	4 (52) 2 (26) UR&RP	9/52 (17%)	5/26 (19%)	5/26 (19%)	14/52 (27%)	5/26 (19%)	5/26 (19%)
Progressive multidimensional	23 (18%)	9 (56%)	4 (52) 1 (13) UR	14/52 (27%)	5/13 (38%)	0	21/52 (40%)	7/13 (54%)	0
Staged conditions	38 (30%)	0 (0%)	4 (52) 1 (13) RP	8/52 (15%)	0	1/13 (8%)	21/52 (40%)	0	1/13 (8%)
Total	125 (100%)	16 (100%)	26 (338) 6(78)UR&RP	58/338 (17%)	12/78 (15%)	15/78 (19%)	127/338 (38%)	24/78 (31%)	19/78 (24%)

EPAR: European Public Assessment Report; N: number; UR: Ultrarare; RP: Repurposed

The level of applicability of the methods differed across the different clusters of conditions.

While all methods were applicable to some extent and in total could add value on average in 76% of the condition clusters, they were often not directly applicable to the actual trial design or approaches used during clinical development of the OMP as described in the EPAR.

The applicability of methods by cluster can be visualized on a heatmap for Step 1: Static analysis (no adjustments) and for Step 2: Dynamic analysis (with adjustments) (Figures 8 and 9).

Figure 8. Heatmap illustrating applicability of methods by cluster *without adjustments*

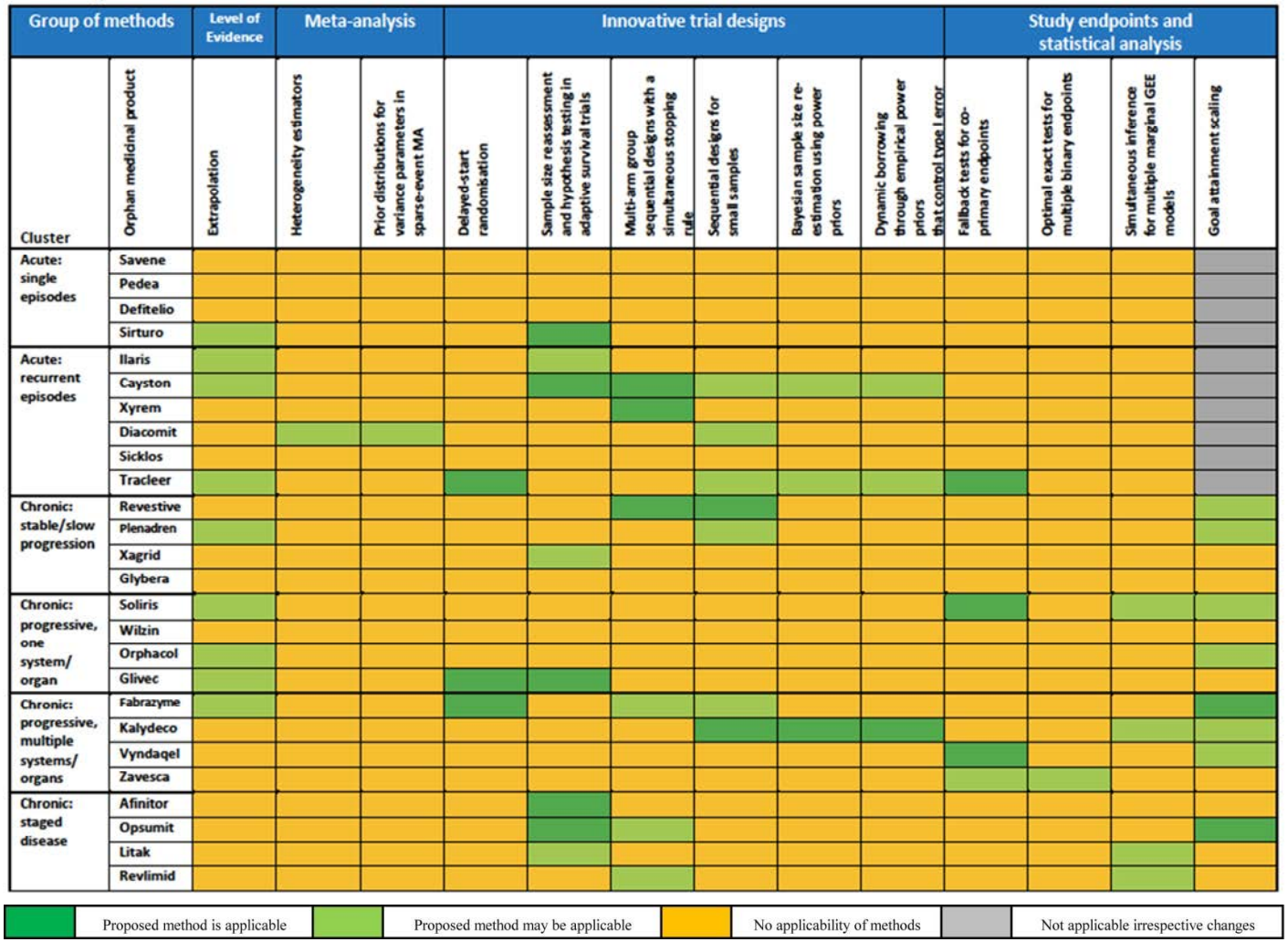
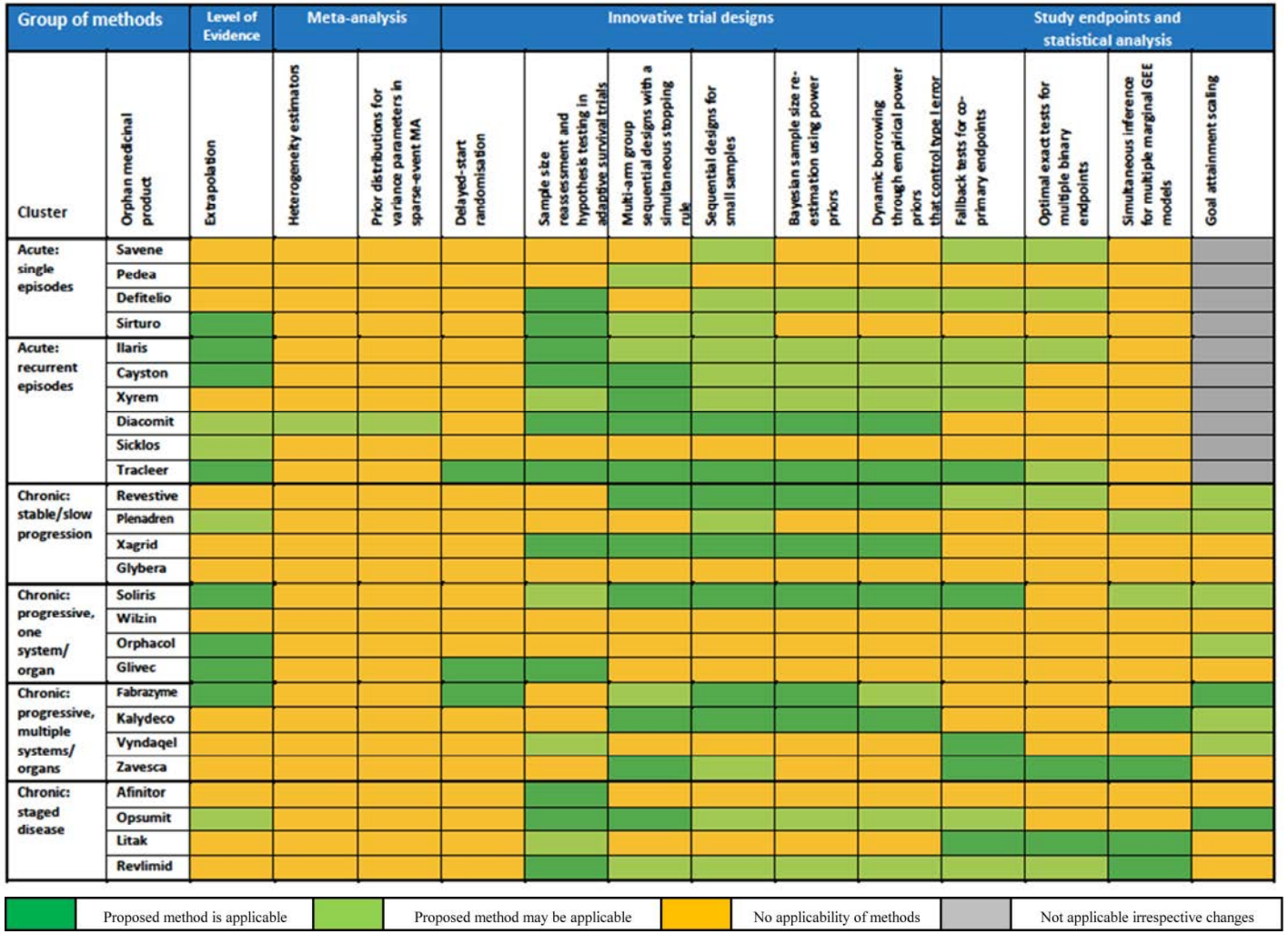


Figure 9. Heatmap illustrating applicability of methods by cluster *with adjustments*



Regarding the applicability by cluster, the findings included the following:

Condition Cluster A: Acute single episodes

For acute single episodes cluster, the requirements for applicability of most designs in this cluster included the characteristics summarized in Table 7.

Table 7. Requirements for applicability in Cluster A: acute single episodes

Actively controlled study in a parallel design
Randomisation
Binary/time-to-event outcomes.
At least 2 experimental treatment arms
Interim analyses
Sample size re-assessments
Prior available data

If the studies were controlled and randomised, had interim analyses and used a binary endpoint or time-to-event measurements, then a number of methods were deemed to be applicable due to the type of endpoints used in this cluster, the limited amount of new potential participant patients and availability of prior data from other similar trials (i.e. for repurposed NSAIDs to treat patent ductus arteriosus).

Contrarily, the methods with limited direct applicability, or no applicability at all, were so because the conditions were acute and so is the expected treatment effect: quick. Therefore, simultaneous inference for multiple marginal GEE models and long-short outcome are not applicable by default from the very beginning. Due to the fact that the diseases within the cluster were a first episode with no previous disease experience, had relatively quickly fatal or disabling outcome and with efficacy measurement based on hard variables, and not at functional level, the GAS cannot be applied either.

Applicability in the cluster is summarized in Table 8.

Table 8. Applicability of methods in Cluster A: acute single episodes

Applicable	Limited direct applicability, or no applicability at all
Sample size reassessment and hypothesis testing in adaptive survival trials (i.e. Sirturo®)	Long-short outcome
Multi-arm multi-stage (MAMS) trial with a simultaneous stopping rule (i.e. Pedea®)	Heterogeneity estimators, prior distributions for variance parameters in sparse-event meta-analysis,
Sequential design for small samples	Delayed-start randomization
Fallback tests for multiple endpoints	Simultaneous inference for multiple marginal GEE models
Optimal exact tests for multiple binary endpoints (i.e. Savene®)	Goal Attainment Scale

Out of the four selected OMP, only one had reported replicated pivotal trials, and it used a binary PEP. However, these were single-arm trials and the expected events were not sparse; hence the heterogeneity estimators and prior distributions for variance parameters in sparse-event meta-analysis were not applicable by default. Due to lack of control, in principle the delayed-start design was not applicable either in the absence of major changes to the clinical development.

Condition Cluster B: Acute recurrent episodes

For Cluster B: acute recurrent episodes, the requirements for applicability of most designs in this cluster included the characteristics summarized in Table 9.

Table 9. Requirements for applicability in Cluster B: acute recurrent episodes

Actively controlled study in a parallel design
Randomisation
Binary/time-to-event outcomes.
At least 2 experimental treatment arms
Interim analyses
Sample size re-assessments
Prior available data
Use of co-primary endpoints to capture the broader array of clinical benefits

A large majority of methods were applicable but not entirely all. Applicability in the cluster is summarized in Table10.

Table 10. Applicability of methods in Cluster B: acute recurrent episodes

Applicable	Limited direct applicability, or no applicability at all
Sample size reassessment and hypothesis testing in adaptive survival trials (for studies supporting Ilaris®, Cayston®, Xyrem® and Tracleer®)	Long-short outcome
MAMS trial with a simultaneous stopping rule for all OMP but Siklos®	Delayed-start randomisation
Sequential design for small samples for all OMP but Siklos®	Simultaneous inference for multiple marginal GEE models
Bayesian sample size re-estimation using power priors and dynamic borrowing through empirical power priors that control type I error methods, using a key secondary endpoint (i.e. FEV1 in Cayston®)	Goal Attainment Scale
Fallback tests for multiple endpoints and optimal exact tests for multiple binary endpoints (i.e. Ilaris®, Cayston® and Xyrem®)	

The sample size reassessment and hypothesis testing in adaptive survival trials (for studies supporting Ilaris®, Cayston®, Xyrem® and Tracleer®) were applicable because they either used a time-to-event endpoint, were or could had been adaptive designs, and either provisioned interim analyses with sample size re-assessment or could had benefitted from this.

MAMS trial with a simultaneous stopping rule was applicable for all OMP but Siklos®, as the studies used a continuous or binary PEP, provisioned interim analyses were designed as sequential trials or could had been designed as such, because of having shorter time-to-outcome as compared to the time needed for enrolment. For the same reasons, and also due to the limited number of patients and due to availability of prior data, sequential design for small samples is applicable for the same reasons as for MAMS. Similarly, these reasons and the use of a key secondary endpoint that was successfully used as primary in other trials (i.e. FEV1 in one study for Cayston®), a Bayesian sample size re-estimation using power priors and dynamic borrowing through empirical power priors that control type I error methods would be applicable, following relevant adjustments.

Fallback tests for multiple endpoints and optimal exact tests for multiple binary endpoints (i.e. Ilaris®, Cayston® and Xyrem®) would be applicable due to the type of endpoints they use, the limited number of patients and availability of prior data from other similar trials.

If the studies were controlled and randomised, provisioned interim analyses and used a binary of time-to-event endpoint, then the above methods regarding study endpoints and statistical analysis would be applicable.

Methods with limited or not applicability at all in this cluster included methods long-short outcome, delayed-start randomisation, and simultaneous inference for multiple marginal GEE models and GAS. Applicability of the methods was restricted by the fact that the conditions

were acute and recurrent in clinical manifestation. Therefore, simultaneous inference for multiple marginal GEE models in repeated measurements and long-short outcome are not applicable by default from the very beginning due to the acute nature. Due to the fact that the diseases are had unstable baseline, relatively quickly fatal or disabling outcome and used hard outcomes, with efficacy measurement not typically at functional level, the GAS would have little or no applicability either.

Out of the six OMP, four reported 2 pivotal trials, one reported one pivotal trial (randomised, placebo controlled) and one was approved based on series of case reports. The complexity and difficulty lied in the fact that out of the OMP with two pivotal trials, only in one case the trials used the same PEP (Diacomit®), for all others the PEP used was different between trials, making the heterogeneity estimators and prior distributions for variance parameters in sparse-event meta-analysis methods not applicable by default.

Condition Cluster C: Chronic stable/slow progression

For Cluster C of chronic stable/slow progression conditions, the requirements for applicability of most designs included the characteristics summarized in Table 11.

Table 11. Requirements for applicability in Cluster C: chronic stable/slow progression conditions

Actively controlled study in a parallel design
Randomisation
Continuous/time-to-event outcomes
At least 2 experimental treatment arms
Interim analyses
Sample size re-assessments
Prior available data
Use of co-primary endpoints to capture the broader array of clinical benefits
Measurements at functional level

A large majority of the methods were applicable but not entirely all. Applicability in the cluster is summarized in Table 12.

Table 12. Applicability of methods in Cluster C: chronic stable/slow progression

Applicable	Limited direct applicability, or no applicability at all
Sample size reassessment and hypothesis testing in adaptive survival trials (Revestive®)	Heterogeneity estimators and prior distributions for variance parameters in sparse-event meta-analysis
MAMS trial with a simultaneous stopping rule (Revestive® and Xagrid®)	Delayed-start design
Sequential designs for small samples (Revestive® and Xagrid®)	Multiple marginal GEE-models
Bayesian sample size re-estimation using power priors, following relevant adjustments	
Dynamic borrowing through empirical power priors that control type I error, following relevant adjustments	
Fallback tests for multiple endpoints (Ilaris®)	
Optimal exact tests for multiple binary endpoints (Ilaris®)	
GAL, if stable baseline and functional measurement	

Examples of applicable methods include sample size reassessment and hypothesis testing in adaptive survival trials in Revestive®, since an adaptive design with provisioned interim analyses and sample size re-assessment was feasible. Similarly, a MAMS trial with a simultaneous stopping rule could be applied to Revestive® and Xagrid®, because the studies could have used a continuous or binary PEP, could have provisioned interim analyses and, instead of four single-arm trials, a randomised trial could have been conducted. Sequential designs for small samples could be applicable for the same reasons as for MAMS, also due to the limited number of patients and to availability of prior data. Thus, the Bayesian sample size re-estimation using power priors and dynamic borrowing through empirical power priors that control type I error methods might also be applicable, following relevant adjustments. Fallback tests for multiple endpoints and optimal exact tests for multiple binary endpoints could be applied to the Ilaris® development, due to the type of endpoints they use, which are not binary but co-primary endpoints. Considering the nature of the conditions in this cluster, a continuous or a time-to-event endpoint may be used and is preferred. If the studies are controlled and randomised, do provision interim analyses and use a continuous or time-to-event endpoint, then the above methods regarding study endpoints and statistical analysis could be applicable.

The methods with limited or no applicability at all included heterogeneity estimators and prior distributions for variance parameters in sparse-event meta-analysis, because the endpoints used were not binary or sparse, the trials used different PEPs or trials were single-arm studies. Lack of randomization and of placebo controls precluded the application of delayed-start design. Since the data was not longitudinally collected, simultaneous inference for multiple marginal

GEE-models was limited in applicability or not applicable. Considering that the diseases exhibit a stable baseline with possible efficacy measurement at functional level, the GAS could be applied.

Out of the four OMP, one reported two pivotal trials, one reported three observational trials following attempt to conduct a randomised controlled trial, one reported a cross-over trial and one was approved based on four single-arm trials following a failed attempt to conduct a randomised controlled trial.

Condition Cluster D: Chronic progressive conditions led by one system/organ

For the cluster of chronic progressive conditions led by one system/organ, the requirements for applicability of most designs included the characteristics summarized in Table 13.

Table 13. Requirements for applicability in Cluster D: Chronic progressive conditions led by one system/organ

Actively controlled study in a parallel design
Randomisation
Continuous/time-to-event outcomes.
At least 2 experimental treatment arms
Interim analyses
Sample size re-assessments
Prior available data
Use of co-primary endpoints to capture the broader array of clinical benefits
Measurements at functional level

A large majority of the methods were applicable but not entirely all. Applicability in the cluster is summarized in Table 14.

Table 14. Applicability of methods in Cluster D: Chronic progressive conditions led by one system/organ

Applicable	Limited direct applicability, or no applicability at all
Sample size reassessment and hypothesis testing in adaptive survival trials (for studies supporting Soliris®)	Heterogeneity estimators and prior distributions for variance parameters in sparse-event meta-analysis
MAMS trial with a simultaneous stopping rule (for Soliris®)	Simultaneous inference for multiple marginal GEE-models
Sequential design for small samples	
Bayesian sample size re-estimation using power priors and dynamic borrowing through empirical power priors that control type I error methods, (Soliris®)	
Fallback tests for multiple endpoints (i.e. Soliris®)	
Delayed-start design, only for Glivec®	
Goal Attainment Scale	

Sample size reassessment and hypothesis testing in adaptive survival trials (for studies supporting Soliris®) are mildly applicable as it could have been an adaptive design with provisioned interim analyses and sample size re-assessment or could have benefitted from this.
MAMS trial with a simultaneous stopping rule (for Soliris®) is considered applicable, as the studies could have used a continuous or binary PEP, provisioned interim analyses. Sequential design for small samples is applicable for the same reasons as for MAMS and also due to the limited number of patients and due to availability of prior data.

For these reasons the Bayesian sample size re-estimation using power priors and dynamic borrowing through empirical power priors that control type I error methods, are or can become applicable following relevant adjustments for the same OMP - Soliris®.

Fallback tests for multiple endpoints (i.e. Soliris®) are applicable due to the use of co-primary endpoints.

If the studies are controlled and randomised, provision interim analyses and use a continuous or time-to-event endpoint then the above methods regarding study endpoints and statistical analysis are applicable.

The methods heterogeneity estimators and prior distributions for variance parameters in sparse-event meta-analysis were not applicable as the endpoints used were not binary or sparse, the trials used different PEPs or the OMP were supported by series of case reports (Orphacol®).

The delayed-start design was applicable only for Glivec® as it is the only scenario where all patients could have been treated eventually with imatinib following a randomised first phase of the trial.

Simultaneous inference for multiple marginal GEE-models was limited in applicability or not applicable due to the fact that the data was not longitudinally collected or due to the series of case reports that supported the applications (i.e. Wilzin® and Orphacol®).

Due to the fact that the diseases exhibit a stable baseline with efficacy measurement at functional level possible but also because the PEP is not entirely representative for all patients, the GAS could be applied.

Out of the four OMP, one reported one randomised controlled trial, one reported a randomised not controlled trial and two did not report any trial nor had for instance a clearly defined PEP. Glivec® is indeed indicated for a rare condition but it is in fact a rare cancer indication presented as an extension of indication after many years in the EU market; and studies for cancer have well-established methodology and considerations, hence there is not much room for adjustments.

Condition cluster E: Chronic progressive conditions led by multiple organs/systems

For chronic progressive conditions led by multiple organs/systems, the requirements for applicability of most designs included the characteristics summarized in Table 15.

Table 15. Requirements for applicability in Cluster E: Chronic progressive conditions led by one system/organ

Actively controlled study in a parallel design
Randomisation
Continuous/binary/time-to-event outcomes.
At least 2 experimental treatment arms
Interim analyses
Sample size re-assessments
Prior available data
Use of co-primary endpoints to capture the broader array of clinical benefits
Short-term outcome predictive for a long-term outcome
Predictable course of disease and not immediately life-threatening for patients
Measurements at functional level
Repeated measurements

For chronic progressive conditions led by multiple organs/systems, a large majority of the methods were applicable but not quite entirely all of them. Applicability in the cluster is summarized in Table 16.

Table 16. Applicability of methods in cluster E: Chronic progressive conditions led by multiple organs/systems

Applicable	Limited direct applicability, or no applicability at all
Sample size reassessment and hypothesis testing in adaptive survival trials (for studies supporting Vyndaqel®)	Heterogeneity estimators and prior distributions for variance parameters in sparse-event meta-analysis
MAMS trial with a simultaneous stopping rule (for Kalydeco® and Zavesca®)	
Sequential design for small samples rule (for Kalydeco® and Zavesca®)	
Bayesian sample size re-estimation using power priors, following relevant adjustments	
Dynamic borrowing through empirical power priors that control type I error methods, following relevant adjustments	
Fallback tests for multiple endpoints (i.e. Vyndaqel® and Zavesca®)	
Optimal exact tests for multiple binary endpoints (i.e. Vyndaqel® and Zavesca®)	
Long-short outcome (Fabrazyme® and Zavesca®)	
Goal attainment Scale	

Examples of applicable methods included sample size reassessment and hypothesis testing in adaptive survival trials (for studies supporting Vyndaqel®).

MAMS trial with a simultaneous stopping rule (for Kalydeco® and Zavesca®), as the studies used a continuous or binary PEP, provisioned interim analyses or could have been designed as sequential due to the shorter time-to-outcome compared to the time needed for enrolment, or they were designed as sequential trials.

Sequential design for small samples is applicable for the same reasons as for MAMS and also due to the limited number of patients and due to availability of prior data.

The Bayesian sample size re-estimation using power priors and dynamic borrowing through empirical power priors that control type I error methods are or can become applicable following relevant adjustments due to availability of data and measured outcomes.

Fallback tests for multiple endpoints and optimal exact tests for multiple binary endpoints (i.e. Vyndaqel® and Zavesca®) due to the co-primary and type of endpoints they use given the multidimensionality of the disease, the limited number of patients and availability of prior data, could be applicable.

The method long-short outcome can be applicable for Fabrazyme® and Zavesca® due to possible use of short-term outcome correlated and predictive for a long-term outcome.

Since the diseases exhibit a stable baseline with efficacy measurement at functional level possible but also because the PEP is not entirely representative for all patients, the GAS could be applied.

If the studies are controlled and randomised, provision interim analyses and use a continuous or time-to-event endpoint, or co-primary endpoints to capture the entire array of clinical benefits in a multidimensional perspective, then the above methods regarding study endpoints and statistical analysis are applicable.

Methods with limited or not applicability at all included heterogeneity estimators and prior distributions for variance parameters in sparse-event meta-analysis were not applicable as the endpoints used were not binary or sparse or the trials were single-arm trials.

Out of the four OMP, two reported a randomised trial, one reported one single-arm pivotal trial, and one reported two pivotal trials.

Condition Cluster F: Chronic staged diseases

For chronic staged conditions, the requirements for applicability of most designs included the characteristics summarized in Table 17.

Table 17. Requirements for applicability in Cluster F: Chronic staged conditions

Actively controlled study in a parallel design
Randomisation
Continuous/ time-to-event outcomes.
At least 2 experimental treatment arms
Interim analyses
Sample size re-assessments
Prior available data
Use of co-primary endpoints to capture the broader array of clinical benefits
Measurements at functional level

For chronic staged conditions, a large majority of the methods were applicable but not entirely all; similar to applicability in cluster E. Applicability is summarized in Table 18.

Table 18. Applicability of methods in Cluster F: Chronic staged diseases

Applicable	Limited direct applicability, or no applicability at all
Long-short outcome (all OMP)	Heterogeneity estimators and prior distributions for variance parameters in sparse-event meta-analysis
Sample size reassessment and hypothesis testing in adaptive survival trials (for all OMP)	
MAMS trial with a simultaneous stopping rule (Opsumit® and Revlimid®)	
Sequential design for small samples (Opsumit® and Revlimid®)	
Bayesian sample size re-estimation using power priors and dynamic borrowing through empirical power priors that control type I error, following adjustments (Opsumit®, Revlimid®)	
Fallback tests for multiple endpoints (i.e. Opsumit®, Litak® and Revlimid®)	
Simultaneous inference for multiple marginal GEE-models (Litak® and Revlimid®)	
Goal Attainment Scale (Opsumit®)	

Examples of applicable methods include the method long-short outcome, which can be applicable for all OMP due to possible use of short-term outcome correlated and predictive for a long-term outcome.

Sample size reassessment and hypothesis testing in adaptive survival trials (for all OMP) are applicable as the studies could have been an adaptive design trial with provisioned interim analyses and sample size re-assessment or could have benefitted from this.

MAMS trial with a simultaneous stopping rule (Opsumit® and Revlimid®) as the studies used or could have used a continuous or binary PEP, provisioned interim analyses could have been applied.

Sequential design for small samples is applicable for the same reasons as for MAMS and also due to the limited number of patients and due to availability of prior data.

For these reasons the Bayesian sample size re-estimation using power priors and dynamic borrowing through empirical power priors that control type I error methods are or can become to some extent applicable following relevant adjustments for Opsumit® and Revlimid®.

Fallback tests for multiple endpoints (i.e. Opsumit®, Litak® and Revlimid®) are applicable due to the use or possible use of co-primary endpoints.

If the studies are controlled and randomised, provision interim analyses and use a continuous or time-to-event endpoint then the above methods, regarding study endpoints and statistical analysis, are applicable.

Methods with limited or not applicability at all included heterogeneity estimators and prior distributions for variance parameters in sparse-event meta-analysis were not applicable as the endpoints used were not binary or sparse or the trials used different PEPs.
Simultaneous inference for multiple marginal GEE-models was limited in applicability due to the fact that the data was longitudinally collected only for Litak® and Revlimid®.

Since the diseases exhibit a stable baseline with efficacy measurement at functional level possible and also because the PEP is not entirely representative for all patients the GAS could be applied (Opsumit®).

Out of the four OMP, two reported one randomised controlled trial, one reported a single-arm trial and one reported a single-arm trial and a randomised controlled trial.

Potential advantages of using new methodology were notably the preservation of operational characteristics, the reduced unnecessary placebo or inferior treatment patient exposure and optimisation of multiple testing procedure; potential disadvantages also were identified such as sufficient level of evidence, but not overwhelming (regardless the positive or detrimental effect on patients.), extra patients needed in case of treatment effect size overestimation and increased complexity or increased resource demand on Sponsors/Investigators.

5.4. Assessment of the regulatory impact

Simulations of alternative clinical development plans for each of the 6 conditions selected are summarized in the upcoming sections.

The results of the simulations of alternative clinical developments for the 6 EPARs (conditions) selected are briefly summarized below. Detailed documents with the full descriptions and analyses conducted for each example, including the summary of evidence and uncertainties, the alternative clinical developments and analyses of the impact of applying these novel methods are presented in Annex 4.

5.4.1. Condition Cluster A: DEFITELIO® (defibrotide) for the treatment of hepatic veno-oclussive disease (VOD)

a) Considerations on the representativeness of the example within the cluster:

- Long-lasting enrolment and important logistic difficulties to the study conduct given that patients are not readily available for enrolment, and due to the high patient dispersion (common within the cluster).

- There are not effective treatments available: due to seriousness of condition, placebo on top of background supportive therapy might in theory be acceptable, but extensive use throughout a previous patient´s named program adds some practical constrains to the placebo use.

- Defibrotide had been in the Italian market for the treatment of peripheral vascular conditions, which gives some supportive safety evidence.

b) Summary of the actual development:

- One pivotal non-randomized (Study 2005-01) vs. historical control in severe VOD. PEP: incidence of complete response by Day 100 post-HSCT.

- Main uncertainties of the actual development plan are related to the lack of an adequate control arm, either concurrent (optimal) or historical (second option). This raises questions on whether the lower rates of mortality observed compared to historical data are just due to improvements in supportive treatment over time, or to differences in the underlying risk of the studied population (87).

c) Summary of alternative developments:

c.1. *Group sequential analysis for small populations.*

A randomized, placebo-controlled trial on top of best supportive care (BSC) is proposed that may address uncertainties of the actual development program by providing randomised double blind evidence on causality. So, it has statistical and regulatory obvious positive consequences, at the price of including and exposing a number of patients to placebo. Although equipoise is evident even now that the product has been authorised, given the high morbid-mortality of the condition this may raise concerns. To minimise this risk, patients will receive BSC, and a sequential design is proposed in order to reduce the sample size to the minimum size required to clear equipoise, and thus, to reduce exposure to placebo and/or to an ineffective treatment.

c.2. S*ample size reassessment and hypothesis testing in adaptive survival studies*

A survival adapted RCT vs. placebo on top of BSC, with 2 interim analyses for futility/superiority/and sample size reassessment, is proposed, which may address the main uncertainties of the actual development program. So, it has statistical and regulatory obvious positive consequences, particularly by including randomisation and placebo control. From a practical point of view, it is less appropriate than other options, since sample size requirements are larger. There may be a reluctance of patients to enter a placebo-controlled trial given that defibrotide is used in clinical practice and due to the high morbid-mortality of the studied condition. The use of background BSC and adaptions of the design based on results of IA may allow early interruption, minimizing exposure to placebo and/or to an ineffective treatment arm, if no longer required to conclude efficacy.

d) Summary of recommendations:

Defitelio in the treatment of VOD associated to HSCT was considered a representative example within the cluster of acute single episode conditions, thus the outcomes of this exercise are deemed generally applicable to the whole cluster of conditions. In particular, the rarity of the condition, the difficulties in recruitment and a short period to response assessment, makes in general adaptions and sequential designs suitable approaches to optimise the trial size that can be generally recommended.

5.4.2. Condition cluster B: ILARIS® (canakinumab) for the prevention of flares in cryopirine-associated periodic syndromes (CAPS)

a) Considerations on the representativeness of the example within the cluster:

- The effect size of canakinumab is outstanding. This makes sample size requirements to test efficacy smaller than in other conditions. Small sample size is likely to be enough to provide reasonable evidence even with conventional designs. In this sense, it may represent an unusual situation.

- No appropriate treatment is available for prevention of flares in CAPS, thus making placebo in theory acceptable. But the biological plausibility is very well established, posing concerns to placebo use.

- Canakinumab had been characterized extensively in other indications providing supportive data to the overall safety and pharmacokinetic development.

b) Summary of the actual development:

- Pivotal evidence was obtained from 1 pivotal randomised withdrawal phase III study (D2304) to demonstrate the efficacy of Ilaris® vs. placebo in patients with Muckle-Wells syndrome (MWS). The study was a 2-arm, 3-stage, double-blind, parallel withdrawal (in stage 2), placebo-controlled study. Thirty-five patients were enrolled in Part I, of whom 31 (responders) were randomized into Part II (15 canakinumab:16 placebo). All these 31 patients entered Part III and received canakinumab. Main endpoint was the proportion of patients with disease flare in Part II (defined as those who experienced a protocol defined clinical relapse or discontinued from Part II for any reason) compared to placebo.

- Main uncertainties relate to the lack of formal dose finding studies, demonstration of efficacy limited to short-term control of inflammation and related symptoms/signs, but the long-term benefits, i.e. reduction/prevention of amyloidosis and end organ damage caused by the inflammatory process, have not been formally demonstrated. Pivotal efficacy from RCT is limited to MWS, while other phenotypes were only included in open label studies.

- The safety database was limited at the time of authorisation: 56 patients had been treated for more than 48 weeks; only 6 patients had been treated for more than 96 weeks. Thus, long term safety beyond 1 year was not well known in this particular population, although it was available for other indications and thus extrapolation was possible (88).

c) Summary of alternative developments:

c.1. *Parallel delayed start pivotal study*

A randomised double-blind delayed start comparison of canakinumab and placebo in the prevention of flares in patients with CAPS is proposed. In stage 1, patients would be randomly assigned to double blind treatment with placebo and canakinumab dosed every 8 weeks (based on phase I PK data). In case of an increase of inflammatory signs or symptoms before 16 weeks patients will be rescued with one dose of canakinumab and switched to second stage. If no response is achieved with rescue, patient is dropped from the study as a failure and treated according to investigator criteria. Second stage will be open-label treatment with canakinumab up to 32 additional weeks (up to 48 weeks). Main efficacy endpoint: proportion of patients with disease flare during stage 1 (defined as those who experienced a protocol-defined clinical relapse, or discontinued study for any reason) compared to placebo.

This option can improve the quality, robustness of the evidence and protection against errors, because it may reduce the placebo effect and any potential effect of the patients knowing that the treatment received is an active intervention, which may overestimate the assessment of the effect through a variable that is clinical in nature. Therefore, a more robust estimation of the effect is anticipated. On the other hand, this alternative option will not improve practical and ethical aspects, in fact, if any, it would have a negative effect as the addition of a placebo arm may delay recruitment and increase exposure to an ineffective treatment. This alternative may be a suitable option and might have facilitated the regulatory access, but it is noted that the main uncertainties would not be addressed. In less severe conditions, such problems are less an issue.

c.2. *Multi-arm group sequential design*

An adaptive design comparing two doses vs. placebo in a sequential manner, with the possibility to drop an ineffective dose for binary variables, is proposed. Interim analyses are planned at 60% and 80% of patients for futility (for the worst dose and for the study), and for overwhelming superiority of any dose. Since in average sequential trials allow for 30% sample size saving, there would be chances to finish with a similar sample size as initially planned, while reducing risk for a negative trial and including assessment of dose response as an additional criterion for concluding on efficacy.

This option may improve credibility in dose-selection, theoretically minimise exposure to ineffective treatments, and reduce time to completion, but given the little room for improvement (due to the overall limited trial size) these are not deemed a major advantage.

However, this could address one of the main uncertainties identified in the development plan conducted, i.e. selection of the optimal dose.

5.4.3. Condition cluster C: REVESTIVE® (teduglutide) for treatment of Short-Bowel Syndrome

a) Considerations on the representativeness of the example within the cluster:

Short bowel syndrome with requirement for parenteral nutrition may have many different causes, have widely variable clinical intensity and affect diverse age groups. Thus, heterogeneity across potential patients may be more important in this particular development than in other medical conditions within the cluster. This may also explain why the clinical program used strict inclusion and exclusion criteria which might limit its external validity, but also, as a consequence, the duration of the trials was longer than expected for a prevalent condition, allowing in this way to consider methods using information as acquired – such methods may not be relevant in other medical conditions within the cluster.

b) Summary of the actual development:

- Two pivotal 24-week double-blind, randomised, placebo-controlled, parallel group studies comparing the efficacy, safety and tolerability of (two doses in one study) of teduglutide and placebo in subjects with Parenteral Nutrition-Dependent Short Bowel Syndrome. The first pivotal trial was non-conclusive, despite it had a major amendment during the study conduction to qualitatively account for heterogeneity, but failed anyway for its main objective for high dose, and thus no further analysis could proceed in a hierarchical testing sequence. Thus, a second pivotal study was done and concluded significant differences vs. placebo. Heterogeneity was lower in this second trial, but external validity was impaired in exchange. The main variable was a response criterion based on a reduction in parenteral nutrition (PN) volume of 20% or more from baseline, or > 2L/week.

- Main uncertainties/weaknesses: The demonstration of efficacy relies on an intermediate variable deemed reasonably representative of the clinical goal, but clinical relevance required expert consensus. There is an issue on the heterogeneity of patients, both regarding the baseline values for volume requirements, severity of clinical impairment, and depending on intake. Heterogeneity was the key determinant in the difficulties of the clinical development to conclude on efficacy of Revestive®. The second study succeeded by reducing heterogeneity, but external validity was impaired in exchange. Risks of the product may include malignancy considering the mechanism of action of teduglutide (growth factor), but the study duration (1 year) is likely unable to detect such risks. The risk/benefit assessment, thus, is uncertain in both the relevance of the clinical changes and the magnitude and severity of potential adverse reactions (89).

c) Summary of alternative developments:

c.1. *Group sequential study with sample size reassessment and simultaneous stopping rule*

A multi-arm group sequential design is proposed as the first (and only, if positive) pivotal trial, with a simultaneous stopping rule using a continuous variable, which could be the change from baseline in PN volume, analysed by Mixed Model for Repeated Measurements test (MMRM) adjusted by baseline value, or the same variable used in the actual development, but untransformed to binary as in the trials (use continuous value for % decrease instead of binary for 20% or more decrease).These changes are deemed as acceptable from a regulatory point of view, and could have allowed the confirmation of efficacy, thus avoiding the need for a second trial. Interim analyses for futility/superiority are planned at 60% of sample size (78 patients) and after 80% of sample size (104 patients). A sample size reassessment is proposed in the second interim analysis as a contingency in case that unexpectedly high variability will occur. Finally, a GAS is introduced as secondary variable in order to support the assessment of clinical relevance of findings.

c.2. *Randomised withdrawal study*

An enrichment design with randomised withdrawal and rescue of failures is proposed, followed by open label long term follow-up period. The design may allow smaller sample size and shorter recruitment, and is likely to observe bigger effect sizes, since all patients will receive active treatment. The ethical assessment suggests that patients may perceive this design as less risky and more open to treatment access. In any case, enhanced access to treatment represents incentives to recruitment that may maximise the practical benefits of the design. Such a design increases efficiency but at the expense of smaller external validity and a potential overestimation of efficacy, which may add to regulatory uncertainties. Besides, the use of GAS provides an intuitive and clinically meaningful sequence of events and allows recommending a therapeutic approach to patient management, although interpretation of results may be hampered by lack of control in the long-term.

5.4.4. Condition cluster D: SOLIRIS® (eculizumab) for paroxysmal nocturnal haemoglobinuria (PNH)

a) Considerations on the representativeness of the example within the cluster:

Some aspects should be considered regarding this example that may differ from other conditions in the cluster:

- The condition is characterized by a long clinical course with an initial impairment of one system/organ, which may involve others along time.

- Progressively reducing life quality and/or quantity of life: typically, subjects are seriously disabled due to disease, but the degree of disability may vary from patient to patient depending on the size of the erythrocyte PNH clone. Due to disease severity, the designs must be add-on to SOC.

- Due to the substantial potential impact on Quality of Life (QoL) of the disease progression, endpoints relying on patient reported outcomes, patient perceptions on the disease, disability and QoL may be relevant or required for decision-making, and thus should be prominently considered in the trials.

- While in the cluster often disease assessment is dependent on patient inputs, with (time to change in) function(s) and QoL being key components of the efficacy measures, this is not the case for complement blockade in PNH. However, another characteristic within the cluster is the assessment of multiple endpoints in the same domain, which is present in Soliris® development.

- Since the condition is chronic with a relatively slow progression, it is acceptable to start research in adults and then progress to the paediatric population.

- Because of the chronic course, recruitment will be based on prevalent cases, but difficulties due to the low prevalence may exist (estimated number of cases of 13 in a million).

- There is a strong scientific rationale for the development of Soliris®, due to the relevant role of the complement in the pathogenesis of the disease. Also, there is a good pharmacodynamic marker (haemolysis parameters including lactate dehydrogenase (LDH)) which may allow for intermediate analysis for the decision making. This may be less established in many other conditions within the cluster.

b) Summary of the actual development:

- Pivotal evidence is supported by one randomised, double blind, placebo-controlled phase III study of 26-week duration that enrolled a total of 87 patients. The co-primary efficacy endpoints for this study were haemoglobin stabilization and units of Packed Red Blood Cells (PRBC) transfused during the treatment phase. Secondary endpoints included transfusion avoidance, haemolysis, and QoL as measured by the FACIT-Fatigue scale.

- The main limitations/uncertainties of the development program of Soliris® in the treatment of PNH are related to the use of short-term surrogate endpoints, i.e. haemolysis, with some supportive clinical outcome data on fatigue, QoL. This was replicated and the effect in the long-term was substantiated by the extension study. The assessment of efficacy in the prevention of thromboembolic events, which is the main cause of mortality in PNH, is based on non-controlled clinical trials. No formal dose-finding studies were conducted. There were also uncertainties related to the characterization of the safety profile of Soliris®, so a safety registry was included as a follow up measure (90).

c) Summary of alternative developments:

c.1. *Multi-arm group sequential designs with a simultaneous stopping rule*

If two different treatment regimens would have been studied (for example in a dose-finding setting) and the two phase 3 studies had merged, then it would have offered the possibility for multi arm multi stage design with a simultaneous stopping rule. Using the simultaneous stopping rule, the critical boundaries are calculated more accurately and using this method, the probability of errantly concluding that one arm is efficacious when in fact it is not, is smaller. Main outcome proposed for this alternative design is haemolysis measured by LDH area under the curve (AUC) during the treatment period. Other endpoints (considered as primary on the actual development) would be considered as secondary. Two Interim analyses are planned at 60% and 80% of the overall sample size using O'Brian-Fleming type boundaries for overwhelming superiority at a two-sided 5% type I error, using multi-arm group sequential design with a simultaneous stopping rule approach.

By merging the two phase 3 studies and applying the multi-arm multi-stage trial with simultaneous stopping rules it can be anticipated some practical and ethical advantages, i.e. reducing sample size, time to completion and facilitating enrolment. The quality of the evidence is not affected in a relevant way and from the regulatory point of view the main advantage would be that this option may address one of the uncertainties of the actually conducted development plan by testing two different dose levels. However, it may negatively affect the ability to provide evidence in clinical outcomes by reducing sample size, shortening duration and relying purely on pharmacodynamics markers. Therefore, this option is not considered to be of true added value, given that no major safety concerns existed on the dose tested and finally recommended for use.

c.2. *Prior distributions for variance parameters in sparse-event meta-analysis*

A prospectively defined meta-analysis of small trials will be planned to generate more robust information on the actual treatment effect on the reduction of the thrombo-embolic events incidence. Thromboembolic events are critical and main cause of mortality associated to PNH. Also, meta-analytic techniques may be applied to the analysis of safety information, considering not only the PNH, but also other indications. Data will be used from the pivotal trial and the open label trial, as well as from the open-label extension studies. It is proposed using two techniques developed within the ASTERIX project, for the primary analysis: the "Prior distributions for variance parameters in sparse-event meta-analysis" and "Heterogeneity estimators in zero cells meta-analysis" as complementary tools.

This approach will allow the fast development based on optimized approaches, with the added value of assessing a hard end-point, to be conducted during the development and likely finalized early after the registration using information from post-authorisation studies. This might eventually be complemented with sequential techniques to further control the type I error and thus be assessed on a sequential basis more than after a fixed amount of information.

Since the estimated prevalence of PNH is only 13 cases per million it will not be expected that the meta-analysis will provide sound confirmatory evidence for the prevention of thromboembolic events, but at least a rough valuable estimate will be available. In fact, this study might provide an overall picture of efficacy and safety assessed in the meta-analysis, and potentially also information in relevant subgroups.

5.4.5. Condition cluster E: FABRAZYME® (agalsidase beta) for Fabry disease

a) Considerations on the representativeness of the example within the cluster*:*

Although considered representative within the cluster, some aspects may differ from other conditions in the cluster:

- Because of slow progression, there are no substantial ethical concerns on starting research in adults before studies are initiated in the paediatric population. Other conditions in this cluster, however, may have rapid progression, or treatment delay may have an impact on prognosis; in such cases, trials including paediatric patients may be justified as the first approach.

- Recruitment will generally be based on prevalent cases and thus already available, so theoretically can be done quickly, unless the condition has very low prevalence and consequently there is wide geographic dispersion of patients (ultra-rare condition, 500-1,000 patients in EU).

- The pathophysiology of the disease (accumulation of sphingolipids due to defective enzyme) is well known and thus the scientific rationale for the development of an enzyme replacement therapy (ERT) is strong. As for similar conditions, this is a chronic life-lasting disease with a highly variable clinical course, with impact in multiple system/organs, requiring multidimensional assessment and endpoints, relying often on subjective assessments from caregivers/patients on clinical or functional status and QoL.

- At the time of the development of Fabrazyme® there were not effective SOC treatments available, so that no active control was identified. Patient management was limited to symptom control and supportive measures.

b) Summary of the actual development:

- Pivotal evidence comes from one phase III randomised, double blind, placebo controlled multi-centre study, conducted in 58 patients, who were randomised to receive 1.0 mg/kg of Fabrazyme® or placebo 1q2w, for up to 20 weeks. The primary efficacy endpoint for this study was the percentage of patients with score 0 on the measurement of reduction of globotriaosylceramide (GL-3) accumulation from the capillary endothelium of the kidney after dosing with randomised study medication for 20 weeks.

Secondary endpoints included other biomarker endpoints, as well as some clinically relevant endpoints.

- Main uncertainties with current development: The demonstration of efficacy was based on pharmacodynamic markers (reduction of sphingolipids in the target organs) with a complete absence of clinical endpoints, i.e. symptoms, function, etc. At the time of decision making for marketing authorisation, there were uncertainties on how changes in sphingolipids may later translate into a clinically relevant effect, and inference on the potential benefit of the product was assumed to derive from the hypothesis and physiopathology. Actually, adverse events (rigors, fever and skeletal pain) were more frequent with active treatment than with placebo, so that in clinical terms and according to available data, the effect could even be deleterious. Such uncertainties were acknowledged, so that the authorisation was issued under exceptional circumstances and a number of post-marketing commitments were requested in order to collect additional long- term efficacy and safety data (91).

c) Summary of alternative developments:

c.1. *Sequential designs for small populations*

If the primary end-point (PEP) were replaced by a continuous endpoint, i.e. based on renal changes such as clearance of creatinine, sequential designs for small populations could be applied. Given the concerns on the sensitivity of the trial to show an effect in creatinine clearance, enrichment approaches might be required focusing on the more severe patients. Sequential designs for small populations may reduce sample size and may shorten time to study completion. Nevertheless, these practical advantages are counterbalanced in this particular case by the fact that patients are already diagnosed and identifiable, and thus available, and at the time the interim analysis is performed, most might have already been enrolled and followed. So, sequential design might delay access to treatment to less severe patients making the general ethical advantages of sequential designs, i.e. minimise exposure to placebo and to the experimental arm, of less value.

Therefore, the usual efficiency of sequential designs is of less relevance in this particular case. By contrary, this design may decrease external validity, the robustness of the evidence provided as it is based on an interim analysis and may also reduce the extent of an already limited safety database.
Furthermore, it might also increase logistical complexity and may delay study completion if new centres were to be opened during the conduct of the trial. For this reason, it is not considered the optimal approach.

c.2. *Multi-arm multi-stage trial with a simultaneous stopping rule.*

Two studies are proposed: first, same as main development for proof of concept and exploration of doses, including 15 subjects to test 3 different doses, two of them with two different schedules, total of 5 treatments, given to 3 patients each. After preliminary

dose selection, a dose selection with pivotal confirmation could follow. This option may improve credibility in dose-selection, theoretically minimise exposure to ineffective treatments, and reduce time to completion, but given the little room for improvement this is not deemed a major added value, in particular since this approach does not address the relevant uncertainties identified in the actual development plan.

c.3. *Methods based on multiple endpoints (i.e. like the Fallback tests for co-primary endpoints and the optimal exact tests for multiple binary endpoints)*, coupled with enrichment methods, where patients with clinical symptoms or functional impairment are included.

Since the disease may affect kidney function, may induce neuropathic pain as one of the key symptoms, has heterogeneity, and has difficulties to measure treatment consequences on health perception, and reasonable dynamic markers can be measured through histology of kidney, skin and heart, there may be room to application of multiple endpoints. Thus, multiple endpoints including clinical outcome measures (symptomatic changes, functional and QoL) and histological changes may provide a more convincing (clinically relevant) demonstration of efficacy.

The advantages of this method are basically that the confirmatory efficiency is increased, as a result of both the likely increase of the effect size by selection of 29/29 patients, and the reduction of risk of failing on the choice of the primary variable, by using a number of related variables with similar and complementary clinical relevance at the same level of confirmatory validity.

Disadvantages may come from the ethical point of view, since the selection of subjects for an enriched design limits the number of subjects who may access the treatment within an experimental setting, and potential delays due to increased recruitment difficulties may lead to longer time to complete the pivotal evidence and thus later regulatory access to the new drug. As regards to the exposure to placebo, the method is neutral, although less patients may be eligible to participate in the trial because of strict inclusion criteria, as already explained. Considering that the condition has quite a slow progression, this may not seem a critical point since there is no expected severe prognostic impact.

In conclusion, the overall balance of an enriched design is a reduction of uncertainty at the price of slower access to active treatment for mildly diseased patients. Using Fallback tests for co-primary endpoints is improving the trial at no substantial impact on other assessment parameters, and thus should be recommended.

5.4.6. Condition cluster F: OPSUMIT® (macitentan) for pulmonary arterial hypertension

a) Considerations on the representativeness of the example within the cluster*:*

- Although Pulmonary Arterial Hypertension (PAH) is a fatal disease, and any treatments should be expected to modify survival, mortality is not usually found in clinical trials

for products intended for PAH as PEP. Alternatively, other endpoints such as PAH related morbidity has been more often used.

- This example that may differ from other conditions in the cluster in that the efficacy of targeting endothelin receptors has been demonstrated previously.

- The development of Opsumit® for the treatment of PAH is characteristic of a non-oncological disease fitting the ASTERIX cluster 6 (staged conditions). There are a number of treatments available for the condition, and there is an available EU regulatory guideline for PAH, acknowledging that it is a disease with very active research. This is not the standard situation in non-oncological orphan conditions, but to note, guidance for oncological conditions (which are most of the conditions included in this cluster) is available, thus the situation can be deemed as representative of the cluster.

b) Summary of the actual development:

- Opsumit® authorisation is supported by a single long term pivotal phase III study (AC-055-302) which is a multicenter, randomised, double blind, placebo controlled study to assess the effects of macitentan on morbidity and mortality in patients with symptomatic pulmonary arterial hypertension.

- Main uncertainties/limitations:

    - Main endpoint of the pivotal trial is a composite variable considering events of different clinical relevance, even if all reflect a negative clinical outcome.

    - The difference detected between the experimental treatments and placebo was already detected after 6 months of follow-up. Longer follow-up was not able to detect effects of macitentan on more finalist variables such as mortality. Thus, it can be questioned the convenience of combining in the same composite endpoint morbidity and mortality.

    - Even if patients assigned to the placebo group were treated with SOC, it can be ethically questioned the value of maintaining this long follow up after the clinical benefit was shown early after the treatment.

    - Finally, no stratification by severity was performed, as defined through WHO-FC, and this would had been desirable to account for clinical heterogeneity (92).

c) Summary of alternative developments:

c.1. *Sample size reassessment and hypothesis testing in adaptive survival studies*

For survival designs, the time to event approach proposed was based on time to first mortality or morbidity event. As some of the clinical morbidity components of the PEP

can be reached early, sequential inspections are regarded as an option to optimize the duration of the trial and/or the number of patients to be included or the assignation ratio (dropping arms). Methods to optimally apply this design to obtain robust evidence have been described within the ASTERIX project.

This method has obvious practical advantages by reducing sample size, exposure to treatments and facilitating recruitment, and thus major advantages from the ethical point of view and for the same reasons. From the regulatory point of view, no added value at all and, if any, some disadvantages as mortality cannot be assessed properly by stopping the trial early and providing less long-term follow-up.

c.2. *Use of co-primary endpoints*

The proposal considers main endpoint of the pivotal trial as composite variable considering events of different clinical relevance even if all reflect a negative clinical outcome. Individual components would be analysed as secondary endpoints hierarchically and with multiplicity adjustments. A different approach based on co-primary endpoints can be explored and methods for the analysis of multiple endpoints as described in the ASTERIX project can be implemented.

The proposal does not represent major changes from the practical and ethical point of view. Some advantages from the regulatory and statistical point of view are obtained, by providing a better protection against type I and II errors, and obtaining results more informative on the clinical relevance, as the impact on each of the clinical endpoints is assessed separately in a non-hierarchical way, which may ease assessment of the benefit/risk balance.

5.5. Final set of basic recommendations as suggestion for implementation in guideline

A summary of recommendations based on the applicability and simulations exercises performed in different conditions for each cluster are presented below.

5.5.1. Condition Cluster A: Acute single episode

This cluster is characterised by incident cases with single acute episodes, with rapid onset and rapid endpoint, well-known and predictable course in absence of treatment, and outcome often serious or life-threatening, with recovery that generally returns to baseline health status with or without sequels. Conditions in this cluster are generally led by one organ/system, although when progressing and with increasing severity multi-organic impairment may occur. The choice of control in the study design must consider whether there is an effective SOC; when a SOC is available, placebo will generally not be acceptable and then comparisons should use add-on designs. In this cluster, efficacy is based generally on a single hard objective and clinically relevant end-point, often binary.

In particular, the rarity of the condition, the difficulties in recruitment and a short period to response assessment, make, in general, adaptions and sequential designs suitable approaches to optimise the trial size. Placebo can be used as control treatment if added to SOC or to BSC.

A sequential design for small populations is a suitable option that may minimise exposure to placebo and optimise trial size. In the simulation it has shown to be more efficient and with less ethical drawbacks than an adaptive survival RCT with sample size reassessment and hypothesis testing, while both improve similarly the robustness of evidence and ability to provide causative evidence.

Bayesian methods may be applied to define stopping rules for both futility and superiority in group sequential analyses. Bayesian approaches to discard minimum required activity to continue an arm are good options for implementation of adaptions if dose finding designs are implemented such as in sequential multi-arm trials with simultaneous stopping rules, or in seamless phase II-III trials joining dose-finding with confirmation of efficacy. Bayesian approaches could be feasible also for implementing sample size reassessment or other adaptions, or for better integration of data increasing information and improving interpretation at the end of any study, especially if information on the condition already exists and is similar to the one obtained in the trial.

For diseases without SOC or trials in patients who have exhausted all SOC options, controls may be historical or external; also, even uncontrolled trials assessing change from baseline or superiority to substantiated expectations may be justified, but then uncertainties are substantial and there are serious difficulties to measure size of effect and to establish causative inferences from data. Thus, in the absence of the possibility to perform a randomized placebo controlled trial due to ethical reasons, the use of high quality historical controls may be an alternative.
Methods for multiple end-points like the Fallback tests for co-primary and the optimal exact test for binary endpoints may not add value unless multiple organs are involved in heterogeneous way.

Conditions with a single acute episode do not allow designs with intrasubject controls, including cross-over and challenge-dechallenge-rechallenge designs.

Longitudinal designs with repeated measures are not suitable for acute single episode conditions.

Delayed start is not a suitable option given the seriousness, usually life threatening, of the condition.

Classic designs remain a gold standard in terms of robustness and would usually be an acceptable option provided that there is effective treatment available or the experimental treatment is given on top of SOC, then allowing placebo control studies.

5.5.2. Condition Cluster B: Acute repeated episodes

This cluster is characterised by prevalent subjects who suffer clear-cut repeated episodes separated by relatively healthy periods. The condition has a well-known predictable clinical course, with repeated clinical episodes led by one organ/system, which are generally due to a single biological or physiological abnormality which -if severe or immunological- may derive into multiorganic impairment. Baseline status may deteriorate along years due to repeated episodes.

In general, there are clinically relevant time-related end-points, measuring the underlying activity of the abnormality through number of episodes by time. If the condition is mild, variables may be based on patient reported outcomes. If the condition is serious, then dichotomic clinical end-points can be used.

Two different indications can be considered: treatment of acute episodes and prevention of new episodes. If the condition is returning to normal after acute episode, start-stop designs (withdrawal, cross-over and intrasubject comparison) may be applicable for both indications. For treatment of acute episodes, variables generally include remission of the episode (binary), time to remission of the episode (time to event) or intensity of the episode. For prevention of the episodes, variables generally rely on number of episodes per time. Longitudinal designs with repeated measures may be applied if the number of events in a period is clinically more important than the time to first episode. Designs where repeated measures are taken into account, with or without weighing by severity of flare, could be an alternative for drugs with lower effect size.

Use of placebo is generally limited in time for non-life-threatening conditions, or when there is lack of prognostic consequences for periods without treatment. A rescue plan is needed (both for placebo or experimental treatment).

Short period to response assessment allow adaptions and sequential approaches in general. A sequential design based on dichotomic assessment of occurrence of first flare may be an efficient approach to this trial. Arm-dropping adaptions may be useful if coupled to sequential approaches to support dose-finding through futility interim assessments – Bayesian methods to discard minimum required activity to continue an arm are good options for implementation of such adaptions.

Bayesian approaches could be feasible also for implementing sample size reassessment or other adaptions, or for better integration of data increasing information and improving interpretation at the end of any study, especially if information on the condition is already existing and similar to the one obtained in the trial.

Tools to ensure that balanced randomization is achieved are important.

Multiple end-points may not add value unless multiple organs are involved in a heterogeneous way.

Enrichment methods are clearly improving ability to detect treatment effect, although may overestimate size of effect and may be less preferable than conventional parallel designs when the latter are feasible.

A remitting – relapsing course is theoretically allowing the use of designs with intrasubject control, the backside being that even if efficacy can be demonstrated, there is less overall subject exposure for safety purposes.

Challenge-dechallenge-rechallenge methods, which were applied in the actual development for the pivotal trial, are particularly compelling and convincing to regulators, since causality conclusions derived from temporal sequence of events are robust and intuitive. Besides, from ethical perspective seem to be almost the only solution in serious conditions with intense clinical expression when a reasonable plausibility on treatment efficacy exists a priori, provided that the response to treatment is not irreversible and follows a close temporal relationship with drug dosing, so that induction of response is quick after drug administration, disease flares back soon after drug withdrawal, and response can be achieved quick again on reintroduction of drug.

External controls, if robust, could provide support to single arm trials, but only if the natural history of the condition is invariably showing poor prognosis and the size of effect in treated subjects is expected to be outstanding. Rates of flares may be compared between groups, or patients may be paired with similar patients in external group, but frequency and type of assessments should be standard enough as to ensure comparability is not impaired.
Delayed start can be a useful approach in this setting, since allows covering ethical concerns (all patients will access the experimental treatment), provides a true baseline assessment, may thus control for overestimation of effect size, and may allow describing prognostic impact and persistence of effect on the long term.

Early rescue shares benefits with delayed start, but if number of failures is anticipated to be high, may compromise blinding and trial integrity, and may difficult the estimation of the effect at the foreseen time of main assessment for efficacy, if substantially later than average rescue.

Classic parallel double blind randomized designs remain a gold standard in terms of robustness, although may be logistically and ethically difficult to implement, and generally require higher sample sizes.

5.5.3. Condition Cluster C: Chronic stable/slow progression

The cluster is characterized by conditions that are life-long lasting and affecting mainly a single system/organ, with constitutive activity due to deficiency or impairment of function and a predictable well-known clinical course. Often there are available surrogates, which measure the underlying defect or function deficiency directly, and often can be validated.

In general, a current SOC or BSC is generally available but not always evidence based. Since prevalence is higher than incidence. Recruitment may be more rapid than subject follow-up, potentially limiting the role for sequential designs and some adaptions.

Because of the relatively stable course of the condition, start– stop based methods (crossover, withdrawal) and methods with intrasubject comparisons may be applicable and useful to control for intersubject heterogeneity.

Given that many outcome measures are based on clinical assessments and may have a subjective component, double blind would be generally required; when standard treatment is well defined, using add-on designs unless treatments share mechanism of action – then direct comparison may be required with non-inferiority approach.

Safety requirements must be widely assessed due to chronicity; especially if conditions are relatively mild with current SOC.

The difficulties for interpretation of clinical relevance and heterogeneity in clinical presentation support that a Goal Attainment Scale may be used as secondary variable to provide support for clinical relevance of findings, especially if methods that allow concluding on small samples, are applied. Due to heterogeneity, stratification by severity is required to *a priori* appropriate management of analysis of subgroups.

5.5.4. Condition Cluster D: Chronic progressive led by one organ/system

The cluster includes conditions with an initial impairment of one system/organ, which may or not involve others along time; clinical course is longer than acute conditions, often year(s). The conditions often are progressively impairing life quality and/or quantity of life, typically subjects are seriously disabled due to disease. Current SOC is generally symptomatic or supportive.

Disease assessment is often highly dependent on patient inputs, with (time to change in) function(s) and QoL being key components of the efficacy measures. Because multiple end-points usually in the same domain may be acceptable/required, co-primary endpoints are useful variables.

There is frequent heterogeneity in clinical expression. Surrogates that allow early (interim) results can be used for decision making but require support for clinical relevance and validation may be difficult if requiring many years and large sample size to conclude on clinical events. Variables are often relying on patient reported outcomes, and patient perceptions on the disease; disability and QoL may be relevant for decision-making. Thus, methods as GAS may be useful to support regulatory assessment of clinical relevance.

Due to progression, start stop methods and intrasubject comparison generally not feasible, and parallel trials are needed when heterogeneity or poor predictability of clinical course are present, with add-on to SOC. Enrichment designs may reduce heterogeneity and sample size requirements, but may impair feasibility of recruitment.
Some adaptions can be applied along the trial. When the condition is severe, classical parallel sequential designs with long term comparison may not be applicable because of ethical concerns and reluctance to randomisation and/or placebo. Early rescue and crossing over of patients after a given period can be useful, such as in delayed start studies. Also, unbalanced randomisation may be useful.

5.5.5. Condition Cluster E: Chronic progressive led by multiple organs/systems

The cluster includes life-lasting diseases, often of paediatric onset, and, if mild or available SOC, affecting (young) adults. Thus, prevalent cases can be expected to be much more frequent than incident cases, and if he condition is not rapidly life-threatening, prospective registries often feasible and available

Parallel designs will be generally needed, due to progression and intersubject variability. Enrichment /stratification may be useful to control heterogeneity. Previous information on the clinical course can be suitable for Bayesian approaches and planning of adaptions. However, sample size adaptions and sequential designs, although applicable, are not considered increasing efficiency if patients are already available for study entry and the use of placebo does not cast major ethical/practical concerns. Comparison on top of SOC allow to manage reluctance to randomisation, but feasibility may be limited if using placebo, especially in paediatric population, or when there is concern on progression and lack of effective SOC. Unbalanced randomisation, delayed start and early escape/crossing over may be useful to limit placebo exposure and cover ethical concerns.

The clinical course may be highly variable, with impact in multiple system/organs that differs across subjects and requiring multidimensional assessment and endpoints.

Variables often are quantitative measurements of impairment of organs or clinical or functional assessments relying on subjective assessments from caregivers/patients. New methodologies taking profit of the multidimensional nature of the condition, like the fallback tests for co-primary endpoints and the optimal exact tests for multiple binary endpoints, are useful options in order to generate a more complete and compelling evidence of efficacy and safety. They may be useful to minimize risk for negative trials due to failure in the choice of the most sensitive variable amongst a number of relevant choices, and also to facilitate generalization of the study results.

QoL is a relevant supportive end-point for regulatory and clinical decision making. Because of heterogeneity and chronicity of functional impairment, the GAS can be useful as a supportive secondary variable to inform on the clinical relevance of findings.
Previous data on event/response rate or variance is often available for current SOC. Many inherited conditions allow development of targeted therapies based on physiopathology. Thus, biomarker surrogates may be useful for early (interim) decision-making, although it may not always be possible to validate along the clinical development. Thus, post-marketing follow-up and registries are a regulatory strategy to handle uncertainties at the time of MAA.

5.5.6. Condition Cluster F: Chronic staged condition

The cluster includes conditions with clearly defined clinical stages which are related to prognosis and difficult to study together. Thus, different severities or extensions of disease have different prognosis and treatment approaches, and disease extension is a key variable, either time dependent or not. Consequently, stratification by the different stages of the condition is often applied to account for heterogeneity when patients in different stages are included in the studies.

Outcomes are generally referred to progression, stagnation or reversal of the condition, with time to change of stage and/or extension of the body or organ involvement as a relevant measure of disease. For those neoplastic, imaging is preferred method for staging; haematological conditions also assess tumour burden, and non-malignant conditions generally measure subject function as a surrogate to organ involvement.

Also, the use of more than one endpoint is an option to obtain a robust evidence of the effect of the treatment. QoL is relevant for all.

Since staged conditions are often chronic or subchronic, a long follow-up is required. The use of sequential methods and adaptive approaches can be useful in this cluster. The use of time to event endpoints is also a characteristic of conditions classified as staged diseases, thus, the proposal of applying methods for sample size reassessment and hypothesis testing in adaptive survival studies is a recommendation that could be valid in many situations.

Stage determines both the design of trials (through stratification of predefined subgroups) and the type of variables (main variable being different in each stage); a variable may be change of status e.g. progression of disease by a given predetermined amount). Also, when feasible, reversal to initial stages or remission of the disease is also a possible endpoint.

If reversal is not feasible, late stages have poor (fatal) prognosis and mortality is a frequent endpoint. Multidimensional and multiple objective measurable end-points would be acceptable in milder conditions; if progression is rapid hard end-points may be feasible as end-points for clinical trials. Often survival designs are applied. Repeated measurements are generally applicable along patient follow-up.

Enrichment designs may use biomarkers selecting potential responders and may reduce sample size requirements but also may difficult recruitment and reduce external validity of results.

Well documented case series on natural course may be available that may allow application of Bayesian approaches and allowing external/historical controls for ultra-rare/poor prognosis.

Because of the progressive course and often the poor prognosis of these conditions, there may be high willingness by patients to accept participation in trials, even when a SOC is available that is not much efficacious; when the prognosis is poor, methods to limit placebo exposure are required to cover ethical concerns, such as unbalanced randomization and early escape or rescue, but frequently trials are single armed.

Safety requirements may be less stringent or delayed if progression is rapid and severe but should always consider impact on QoL.

6. DISCUSSION

6.1. Regulatory uncertainties associated to OMP development

Efficacy of new treatments requires scientific demonstration before such treatments are deemed appropriate for use in routine clinical practice. Furthermore, the risks associated to the use of new medicinal products require qualification and quantification to ensure that any potential impact on public health can be foreseen and managed. The analysis of the benefits and risks balance of a new treatment is the basis for the regulatory assessment of new products, and thus requires a detailed description of both anticipated benefits and risks. This analysis takes into account not only the demonstrated benefits and risks, but also the existing uncertainties and limitations of the evidence on benefits and risks (93). The higher the level of uncertainties at the time of making this analysis, the less solid and robust the basis for a conclusion is, with the potential negative consequences that weak decisions may have for the public health (94–96). Therefore, well designed and sufficiently powered clinical experiments are the way to collect robust efficacy and safety information.

However, generating robust evidence with small subject samples is a methodological and logistic challenge that may discourage sponsors from researching new treatments for rare diseases. The poor understanding of the natural history of the disease progression, difficulties to achieve an accurate diagnosis, heterogeneous patient populations with variable phenotypes and clinical courses, limited number of patients plus geographic dispersion of patients and investigators, regulatory uncertainties, difficulties in choosing clinically relevant outcomes, and lack of prior experience in conducting clinical studies that could serve as guidance, have all been highlighted among the main difficulties (7,52).

These particularities explain why orphan drugs development pose unique challenges and, in spite of the advances made during the past years, recent analyses conclude that difficulties in orphan drug development persist (29).

As a consequence, the regulatory decision for OMP has to be taken with higher uncertainties than those handled in conventional diseases. This has been reflected by several authors.

Picavet at al. found that the pivotal studies that are the basis for marketing authorization of OMP exhibit methodological flaws that are a cause of important concern i.e. the lack of QoL-related endpoints as outcome, lack of blinding in the study design and the use of surrogate endpoints, and concluded that a more demanding regulatory process for OMP is needed to guide evidence-based clinical decision-making (97).

Dupont et al. analysed the Belgian reimbursement decisions of orphan drugs as compared with those of innovative drugs for more common but equally severe diseases. The authors found that only 52% of the 25 orphan drug files included a RCT as opposed to 84% in a random control sample of 25 non-orphan innovative submissions ($P < 0.01$). Despite the poorer quality of the evidence supporting OMP marketing authorisations, the proportion of submissions that were granted reimbursement was statistically significantly higher for orphan than for non-orphan innovative medicines (88% vs. 63%, respectively, $p= 0.02$) (98).

Another comparative analysis of the characteristics of pivotal trials to support FDA approval of orphan vs. non-orphan drugs in cancer was conducted by Kesselheim A. et al. The authors concluded that, compared with pivotal trials used to approve non-orphan cancer drugs, pivotal trials for recently approved orphan drugs for cancer were more likely to be smaller and to use non-randomized, unblinded trial designs and surrogate end points to assess efficacy (16).

A similar exercise was performed by Mitsumoto J. et al., who found that while all drugs for neurological diseases approved without an orphan indication included at least two randomized, double-blind, placebo-controlled trials, by contrary, 32% of drugs with an orphan indication had at least two such trials ($p < 0.001$), 74% had at least one ($p = 0.02$), 33% did not use a placebo control, 27% were not double blind, and 12% were not randomised (17).

Consistent results were found by Bell SA. et al. who compared the characteristics of interventional clinical trials registered at ClinicalTrials.gov in rare versus non-rare diseases up to September 2012, and found that rare disease interventional trials differ substantially from those in non-rare conditions, with differences in enrolment, design, blinding and randomization: rare disease trials enrolled fewer participants (median 29 vs. 62), were more likely to be single arm (63.0% vs. 29.6%), non-randomised (64.5% vs. 36.1%) and open label(78.7% vs. 52.2%) (15).

More recently, an analysis of the regulatory evidence supporting EU OMP authorization from the start of the Orphan Regulation (EC) 141/2000 until December 2014 concluded that the regulatory evidence supporting these authorizations showed substantial uncertainties, including weak protection against errors, substantial use of designs unsuited for conclusions on causality, use of intermediate variables without validation, lack of apriorism and insufficient safety data to quantify risks of a relevant magnitude (42).

In this context, the already existing and newly developed regulatory tools aimed to foster early access to innovative medicinal products in areas of high unmet medical need (which is usually the case for most OD) add further uncertainties to the regulatory decision-making process. Regulatory provisions, like the conditional marketing authorization (41) and initiatives for the adaptive licensing (99), substantially increase the level of uncertainties at the time of drawing a regulatory opinion, as long as these tools allow concluding on the basis of less amount of information or an incomplete clinical data package, i.e. before comprehensive data are available.

The attempts to ensure a timely access to innovative medicines in areas of unmet medical need pose additional risks, as long as uncertainties faced at the time of granting a marketing authorisation are expected to be duly and timely addressed during the post-marketing. However, some published data indicate that this is not usually the case. A study conducted by Hatswell, A J. et al. revealed that, in spite of the frequency with which approvals are granted without RCT results, there is no systematic monitoring of such treatments to confirm their effectiveness or consistency regarding when this form of non-randomised evidence is appropriate(100). Furthermore, a systematic evaluation of oncology drugs approvals by the EMA showed that most cancer drugs entered the market without evidence of benefit on survival or quality of life, and that after a minimum of 3.3 years from market entry, there was still no conclusive evidence that these drugs either extended or improved life for most cancer indications (101).

Therefore, OMP marketing authorizations face a high-level of uncertainty during the decision-making process, and this may have obvious negative consequences for the public health. In addition to the potential misuse/waste of limited economical health resources that this may imply, uncertainties on the actual benefits and risks of a given medicinal product may expose patients to potentially unacceptable safety risks, and/or to the risk of losing a window of therapeutic opportunity by receiving a potentially ineffective drug (95,96,102).

Thus, there is a huge and urgent opportunity for improvement, both for making a more efficient use of available human and economic resources, and for generating sufficiently robust evidence to allow for better benefit risk balance assessment with a substantially lower level of uncertainties than currently faced. This is particularly important in front of the challenges posed by the early access initiatives.

6.2. The role of novel methodologies in orphan drugs developments

Many barriers exist to advancing knowledge of, and finding treatment options for, rare diseases. The small target populations can dampen commercial interest in development of treatments. But even for those rare conditions where manufacturers of therapeutics are engaged, methodological and data constraints limit the ability to generate evidence on patient health outcomes (52).

There are no special methods for the design, the operational execution or the analysis of clinical trials in small populations. Conventional parallel group RCT, which randomly allocate participants to one of two or more treatment groups, are not always feasible in rare conditions. However, in recent years, innovative epidemiological and clinical trial methods have been developed which may constitute viable pathways for collectively advancing in the study of rare diseases. These novel methodologies offer promise for promoting more efficient and effective research.

Gagne J.J. at al. conducted a methodological review to catalogue and describe innovative approaches to studying health outcomes in patients with rare diseases, that have been, or can be, applied to overcome the methodological challenges inherent to the study of rare diseases (52). Designs aimed to reduce sample size requirements include adjustments to traditional randomized trials like, e.g. choosing longer trial duration, use enriched populations, or tackling multiple treatment options in a factorial study. Also, selection of outcome measures more sensitive to changes, like a continuous outcome variable, a surrogate marker, a composite endpoint, or repeated measure outcomes, may all allow capturing outcomes more precisely, or more events among the trial participants, thus reducing sample size requirements. In addition, novel trial design strategies suitable to account for small pools of patients with rare diseases have also been proposed, like trials featuring flexible designs that allow modification of some aspects of the trial based on prospectively planned interim data analyses, e.g., adaptive randomization and sequential trials, and its many variants. Despite using these methods, individual trials may remain underpowered. To overcome this problem, results can be incorporated into a prospectively planned meta-analysis or into a Bayesian framework.

There are also strategies aimed to maximize information obtained from on-treatment participants, to facilitate recruitment for patients with rare diseases who have limited treatment options, and, at the same time, some designs can also reduce recruitment requirements, e.g. crossover trials and its variations like the randomized withdrawal studies.

Further, new methods and refinement of previously described alternative designs have been developed in the recent years (20,71–83). Alternative methods are different methodological approaches aimed to increase the efficiency of clinical trials, and thus can help in addressing many of the challenges posed by clinical developments of orphan drugs. Paradoxically, they have been mostly used in large clinical trials supporting marketing authorisation decisions for prevalent conditions.

Therefore, there is need to increase awareness of the armamentarium of available research tools and their utility. Furthermore, there is also a need to continue developing new designs and statistical methods, or refining the existing ones, in order to face properly the challenges of orphan medicines development. This will contribute to the conduct of more efficient clinical development plans for OMP and to generate more robust evidence, able to substantiate more solid benefit/risk decisions.

6.3. Regulatory tools aimed to assist applicants during OMP development

Even with all the evidence available, regulatory uncertainties during the decision-making process are commonly present. In example, there may be uncertainty in benefit stemming from limits in our scientific understanding of a disease, from inconsistent or contradictory evidence across multiple studies, or from the relationship between study population and those who will actually take the drug. Uncertainty about risks may stem from numerical imbalances of adverse events in treatment and control groups where statistical testing is not reliable due to lack of pre-determination, small size and multiplicity concerns, or from post-market data obtained from sources of varying levels of rigor, or from the ability of the health care system to adequately manage a risky drug.

In more complex situations, like is usually the case for orphan drug developments, uncertainties can be significant. These are derived not only from the logistic and methodological challenges faced during drug development and the necessarily reduced evidence-base for the decision-making, but also for the increasing frequency in which decisions are taken on the basis of an immature or incomplete dossier, i.e. when early access tools are accessed (103).

Drawing conclusions in the face of uncertainty is a complex task, where subjective value judgements take a relevant role, thus increasing the likelihood of wrong decisions. This poses important risks, mostly derived from the potential consequences for the public health that may have to grant a marketing authorisation for a potentially ineffective treatment and/or for a drug with a worse than expected and hardly acceptable safety profile (102).

Reducing uncertainties at the time of granting a marketing authorisation is critical to facilitate the decision-making process and to minimize any potential risks for the public health.

Thus, regulators are particularly committed to support developers on the appropriate studies to be conducted in the development of a medicine in order to generate adequate data for the benefit-risk assessment at the time of MAA, and thereby facilitating the introduction of new, safe and effective medicines (24).

In the particular case of OMP development, the lack of prior experience in the conduction of trials in OD to guide developments, and also the absence of specific regulatory guidelines, causes uncertainty for drug developers on what to provide with the approval package, particularly when not prior experience does exist.

Therefore, a more thorough assistance to the sponsors early during clinical drug development is needed. Scientific advice procedure can help in addressing challenges for individual cases and increase the likelihood of a favorable outcome as shown by Regnstrom J et al. and Hofer M et al. (24,40,104). However, in view of the repeated profile of limitations of orphan drug developments, and the frequently asked questions and requests for scientific advice procedures, it is clear that there is a need to strengthen the methodological recommendations applicable to the development of OMP by means of regulatory guidelines. This will likely help to address the methodological flaws often encountered in orphan drug developments.

6.4. The ASTERIX Project

The objectives of the ASTERIX Project (55) were to develop design and analysis methods for single trials and series of trials in small populations, to assess the applicability of these new methods to orphan conditions, and to provide a set of recommendations. The ASTERIX Project was a unique effort aimed to address the challenges of OMP developments by developing novel methods and designs that might be more efficient or robust in small populations, or better suited for these conditions, thus addressing the methodological flaws of OMP developments. This project was ultimately aimed to provide a practical guidance on the applicability and added value of these methods to the different orphan conditions, in order to assist Applicants during the planning and developing process of an OMP, and consequently, to facilitate the regulatory decision-making process.

Different frameworks are available that are typically aimed to propose algorithms or decision processes to arrive at the most suited design for a given clinical trial. These frameworks focus either on a specific condition (105,106), a specific method or group of methods (107,108), or provide general recommendations (109–112). Some of the proposed frameworks have attempted to tie attributes of interventions and rare diseases to specific methodological approaches.

Cornu C. et al. proposed an algorithm for the choice of an appropriate trial design in the development of orphan drugs based on a number of identified design characteristics that seemed most likely to guide the choice of a specific design. These included the use of reversible or irreversible outcomes, fast (defined as up to a few weeks) or slow response to treatment, possibility of minimising the time on placebo, possibility that all patients received active treatment by the end of the trial, and possibility of performing intra-patient or inter-patient comparisons (109).

Gupta and colleagues proposed an alternative framework to help investigators to determine when different designs are appropriate. The framework takes investigators through a series of questions to assess the usefulness of alternative designs in particular situations. The choice between methods is guided by factors related to the intervention, disease, anticipated recruitment duration and success, and current state of knowledge about the treatment (110).

A different framework was proposed by Parmar M. at al., aimed for designing randomised trials and to address the problem when the ideal sample size is considered larger than the number of participants that can be recruited in a reasonable time frame. Staying with the frequentist approach, the authors propose a framework that includes small iterative alterations to the design parameters with the aim to increase the numbers achievable, and also potentially reduce the sample size target (111).

Abrahamyan et al. presented a conceptual framework to guide researchers in choosing among a number of study designs based on the characteristics of the intervention (if it has a predictable and rapidly reversible effect), if the time between study inclusion and outcome assessment is short compared with accrual time, if therapy is likely to provide a lasting response, and the possibility of randomization to placebo (112).

Recently, Whiecher D. et al. (113) reviewed all these available algorithms for matching study design to rare disease characteristics, and summarized applicable methodological and analytic approaches. The authors conclude that using these approaches can facilitate the completion of RCTs that are adequately powered and proposed the creation of an effective research infrastructure that could help in prioritizing studies, accelerate accrual, catalyze patient engagement, and avoid waste in research.

However, most of these general algorithms or frameworks are guided by items related to only a few characteristics of the condition such as clinical course, timing and reversibility of the outcome, or trial feasibility, and they are not always exhaustive to fit all possible situations. Furthermore, these frameworks in general only account for the frequentist methods, which are more familiar to researchers, review bodies and regulators, keeping Bayesian approaches and novel methodologies recently developed aside.

Importantly, all these algorithms focus on the design of the pivotal studies but *do not consider the development program in its totality*, and how novel designs and methods of analysis can contribute to an efficient use of resources while generating the needed solid evidence to substantiate regulatory and clinical practice decisions. Since early decisions in the clinical development process are increasingly more important as early access is a reality, adequate time should be set aside and early in the development to carefully plan the overall development plan and get the design of the clinical studies right (111).

The limitations of the existing frameworks to provide guidance that directly incorporates characteristics of the medical condition treated are obvious. Apart from their low prevalence, OD is a highly heterogeneous group of diseases. Such heterogeneity makes it very difficult to issue useful regulatory recommendations relevant to all (or at least most) possible clinical situations in the course of uncommon diseases. Nevertheless, some groups of conditions share similar clinical characteristics linked to the applicability of certain trial designs and general

approaches (61). Thus, within the ASTERIX project, a heuristic framework was proposed that could help identify groups of medical conditions for which similar methods could be useful for drug development ((114) to appear), (115).

The framework of medical condition clusters is considered a practical tool to guide developers throughout the different methods and design analysis. The advantage of this approach is that guidance is provided taking the main characteristics of the disease course and the intended therapeutic indication as the starting point, which is a similar approach to that followed when planning the developing program of a given medicinal product. Thus, from a regulatory point of view it is considered an intuitive and practical approach. Furthermore, going from the disease context and the intended therapeutic goal is also the way clinical investigators face research questions aimed to address clinical practice problems. Therefore, taking into account the key role that academia research plays in orphan drug developments, the cluster of medical conditions is also considered an intuitive approach matching the researcher´s point of view.

For these reasons, the proposed cluster of medical conditions was applied as a way to structure the evaluation of the applicability of (novel) methods and their added value, and is considered a suitable tool to provide guidance, which could be given more specifically at the condition cluster level.

In that sense, the present approach can be considered similar to some of the existing algorithms. However, the current proposal goes beyond these approaches, which just focus in choosing the best design for the pivotal study and considers the applicability of all existing methods and their added value in the context of a complete development program.

6.5. Revision of available EU regulatory guidelines relevant for orphan diseases

The lack of prior experience in the conduct of trials that could guide developments and the absence of specific regulatory guidelines causes uncertainty for drug developers on what to provide with the approval package, particularly when not prior experience does exist.

The systematic search conducted for EU guidelines related to the conduct of clinical trials of an OMP development plan was firstly intended to establish the actual situation with regards to availability of regulatory guidance for orphan conditions. It was found that 39% (71/182) of EMA's guidance documents were applicable to the clinical development of OMP; 25.3% (46/182) of EMA´s documents had general applicability, 12.6% (23/182) had specific applicability, and 1.1% (2/182) both general applicability as well as specificity was revealed. Formal reference to rare diseases and/or OMP were found in up to 50.7% (36 of the 71) of documents considered applicable, but it is noted that 14.1% of documents (10 out of the 36 documents with formal reference) the reference to rare diseases was accompanied with limited or none specific information on the characteristics which are considered important when designing and conducting a clinical trial.

These results revealed the existence of some but limited and disperse guidance to provide references on the design of studies for the demonstration of efficacy (and safety) of medicinal products in small populations suffering from a rare disease. Less than half of the existing regulatory documents are applicable to OMP investigations, while those specific to rare diseases are even fewer.

Among the documents with general applicability, those that had references to rare diseases were found to be less than one quarter. This could be explained as if the same principle would apply regardless of the prevalence of the condition or, by contrary, that orphan conditions are just the anticipated exceptions where seeking scientific advice is recommended.

On the other hand, the experience gained during these past years has led to the development of disease specific guidelines for some orphan conditions, where more specific and concise recommendations were found in line with those existing for more prevalent conditions. Nevertheless, it was found that often similar recommendations with respect to requirements for the design of main studies are given for substantially different diseases that have similar course of disease, or certain medical characteristics. In other cases, information was found to be not so concise regarding the rare conditions, e.g. the case of other dementias that were included in the guideline dealing with Alzheimer disease.

Consistently, most of documents specifically applicable to OMP investigation included very scarce recommendations on statistics. In the most recent guidelines, cross-reference to the specific biostatistics guidelines was nevertheless found.

Considering all the above, the information that was found to be specifically addressed to clinical trials for OMP investigation, is considered fragmented and not always consistent.

As this review revealed, general applicability guidelines, even if mentioning rare diseases, do not cover most of the issues relevant to the development of OMP, such as the level of evidence needed or the acceptability of innovative methods. In example, although the value of novel methodologies to the design and analysis of studies in small populations is recognized, the only document mentioning them is the Guideline on clinical trials in small populations(49), which summarises a range of possible approaches in the context of small populations in drug development, acknowledging that any efficiency improvements for small population clinical trials would also be relevant to larger trials and vice-versa.

Therefore, identifying aspects that are more controversial when dealing with rare diseases and developing specific guidance on these issues were deemed a good starting point in order to enhance the clinical investigation of OMP. This will likely help to address the methodological flaws often encountered in orphan drug developments. Nevertheless, it is acknowledged that more specific guidance is considered difficult at disease level due to different limitations; including the large number of identified OD most of them with scarce knowledge on its natural history, deficient diagnostic methods and non-existent previous drug development. Our alternative approach, whereby recommendations could be given at a group of conditions level, appeared a more realistic one.

6.6. Evaluation of the applicability of the methods developed within ASTERIX Project

The new methods developed in ASTERIX included new proposals for interim analysis and stopping rules in multi-arm parallel trials, methods for sample-size reassessment, rules to optimise boundaries in group-sequential designs, methods to tune the use of prior information from similar trials in Bayesian analysis, considerations to apply flexibility to the level of evidence, new approaches to analyse multiple endpoints, a patient-centered instrument for heterogeneous functional outcomes and two methods for meta-analysis of sparse binary data.

The applicability requirements for the methods included mainly the type of measurement (i.e., binary or continuous variable, single or multiple main endpoint, scarcity of data), availability of more than one trial, availability of previous studies with good quality data, the length of time to end-point as compared to the time to complete recruitment, and feasibility of randomized designs.

The potential applicability of methods and advantages was evaluated based on information from actual trials. In the first, static step, it was found that all individual methods were directly applicable to a minimum of 1 (4%) up to 9 (35%) of the 26 EPARs, and overall each method was applicable to a minimum of 1 (17%) and a maximum of 5 (83%) of the 6 clusters. In the second, dynamic step, it was found that the individual methods were applicable in 1 (4%) up to 17 (65%) of the EPARs, and a minimum of 1 (17%) out of 6 and a maximum of 6 (100%) of the 6 clusters.

In general, the applicability of methods within the disease clusters can be summarised as follows:

- Condition cluster A: acute single episodes can benefit most from innovative trial designs and partially from methods addressing the study endpoints. GAS cannot be applied. Overall, up to 27% of the tests concluded some degree of applicability for the ASTERIX novel methods.

- Condition cluster B: acute recurrent episodes, was a good ground for all groups of methods, especially for innovative trial designs and also for meta-analysis methods. GAS cannot be applied. Overall, up to 53% of the tests concluded some degree of applicability for the ASTERIX novel methods.

- Condition cluster C: chronic stable/slow progression contained OMP for which all methods groups were applicable, including GAS, except for meta-analysis methods. Overall, up to 31% of the tests concluded some degree of applicability for the ASTERIX novel methods.

- Condition cluster D: chronic progressive led by one system/organ, contained OMP for which all methods were mildly applicable to some extent, except for meta-analysis methods. GAS can be applied in this cluster. Overall, up to 27% of the tests concluded some degree of applicability for the ASTERIX novel methods.

- Condition cluster E: chronic progressive led by multiple systems/organs, contained OMP for which all methods groups were applicable, except for meta-analysis methods. For this cluster, GAS was found particularly applicable. Overall, up to 40% of the tests concluded some degree of applicability for the ASTERIX novel methods.

- Condition cluster F: chronic staged diseases contained OMP for which all methods groups were applicable, except for meta-analysis methods. For this cluster, GAS was also found applicable. Overall, up to 40% of the tests concluded some degree of applicability for the ASTERIX novel methods.

While all ASTERIX methods evaluated were applicable to some extent in at least one situation, and in total could add value on average in 76% of the condition clusters, they were often not directly applicable to the actual trial design or approaches used during clinical development of the OMP as described in the EPAR.

Previously proposed frameworks for guiding on the applicability of novel methodologies relied on a single characteristic of the condition under investigation to determine the suitable methods. However, as recognized by Abrahamyan at al.(116), recommendations guided by a single characteristic might not be a useful one given that many of the characteristics relevant to determine the applicability of methods are not mutually exclusive and thus, several approaches might be feasible for a single research question. Thus, for the first time the applicability exercise was performed in a limited group of conditions that are defined by some of the most relevant aspects that determine general applicability of methods. This allow considering all the possibilities applicable in a particular case, and to select those finally considered most suitable to address the challenges of this particular case.

Nevertheless, this was not an exhaustive exercise, so that applicability within the clusters was based on the 4-6 EPARs evaluated within each cluster and thus might not be fully generalizable to all conditions and drug development plans within the cluster. Further, given that the total number of EPARs within each cluster varied substantially (ranging from 9 to 38 in cluster B acute recurrent episodes and in cluster F chronic staged conditions, respectively), the quantitative representativeness of the examples within the clusters differed substantially (ranging from 10% of all existing EPARs in cluster F chronic staged conditions being tested up to 67% of all existing EPARs being tested for cluster B acute recurrent episodes, respectively). This has obvious consequences for the quantitative conclusions on the general applicability of methods and their generalizability within the cluster.

The value of this exercise is that it gives valuable information on the general applicability of the novel methods developed within the ASTERIX Project. Further, the exercise in itself also showed that a systematic approach including the definition of the applicability pre-requisites of the methods, together with the definition of the general characteristics of the medical conditions included in a given cluster, allows guidance to investigators on whether they could consider a given method or not for a certain type of medical conditions, which is considered a useful practical approach.

On the other hand, the overall limited applicability of the methods in this exercise is not totally unexpected considering that the focus was solely placed on the pivotal studies of the actually conducted development programs, and all considered alternatives kept the primary development objectives intact.

We may hypothesize that the situation might differ when the focus is extended to the complete development program, and consideration to the novel methodologies is carefully given early in clinical development. This was the rationale behind the additional simulations exercise.

6.7. Simulation of clinical developments applying novel methods to real examples

Initially we aimed to validate the applicability and added value of novel methods and approaches by testing these methods in raw data from selected clinical trials. However, access to real-life raw data happened to be a long-lasting bureaucratic procedure that made it unfeasible granting access within the required tight timelines for the ASTERIX Project completeness. Instead, we decided to simulate as close as possible the real situation of planning a clinical development program. Therefore, information from the EPARs was used as a contingency and allowed doing a qualitative analysis to judge the added value of applying novel methodologies. This approach might be less suitable to properly assess the added value in terms of robustness of data, but instead the overall utility from the regulatory point of view can be better estimated.

Simulations were performed to check the alternatives and gains for six real examples belonging to different clusters of medical conditions. So, for a total 6 conditions selected from each cluster alternative development plans to the one actually conducted were simulated, applying (novel) methods based on previous conclusions of applicability. This qualitative exercise was aimed as an attempt to address the regulatory uncertainties identified by the CHMP at the time of drawing an opinion on the marketing authorisation by means of the (novel) methods. The impact on other regulatory, practical, ethical and/or statistical aspects were systematically assessed to ensure that addressing the regulatory uncertainties was not done at the cost of severely impacting any of these relevant aspects for orphan conditions.

6.7.1 Defitelio®

For Defitelio® (defibrotide), an example from Cluster A: acute single episode conditions, the main uncertainties of the actual development plan were related to the lack of an adequate control arm, making this a weak basis for decision making on the risk/benefit.

Previously to the design of the pivotal study, there were data from 6 case series, all of them gathering data from compassionate use, and a randomised open-labelled trial comparing two dosage schedules of Defitelio®. Responses varied between 36% to 100% and survival at 100 days between 32% and 93% (117). At the time of designing the pivotal clinical study there was a high expectancy of efficacy, and access to the treatment was granted via compassionate use programs, thus, making the feasibility for conducting RCT highly questionable.

To note, the efficacy rates obtained from previous studies showed very variable rates for both response and survival at 100 days, and the dose selection was quite uncertain. In the absence of an adequate control arm for the pivotal study, either concurrent or historical, uncertainties related to the actual treatment effect and the added toxicity of defibrotide in the context of a high morbi-mortality disorder could not be easily addressed.

These uncertainties have serious implications at different levels. First, there were important doubts on the appropriateness of the level of evidence for the benefit risk balance assessment of Defitelio®. Secondly, these uncertainties were unlikely to be solved in the future since randomised controlled trials would become even more unfeasible once the marketing authorisation was granted. Finally, orphan market exclusivity would prevent a fair scientific competition in case of better and more robust trial designs for potential competitors, given that any development program will obviously require the demonstration of either non-inferiority or superiority designs for competitors. The highly variable efficacy rates and the lack of placebo-controlled information will made impracticable the estimation of an appropriate non-inferiority margin and the likely overestimated efficacy rates will difficult the design and feasibility of superiority trials compared to Defitelio®. All these issues are well known and they have been pointed out and discussed in the literature in the context of a convincing example(20). The key conclusion is that the development plan must always be initiated with randomised designs, and this is critical in the case of small populations. Once (likely biased and overestimated(118–120)) efficacy data is obtained from uncontrolled designs, is practically impossible to promote controlled design in the field of small populations (20).

These main uncertainties could have been addressed by providing randomised double-blind placebo-controlled evidence on causality by means, in example, of any of the two (novel) designs proposed, i.e. a randomized placebo-controlled trial over SOC with sequential analysis for small populations, or an RCT trial with sample size reassessment and hypothesis testing in adaptive survival studies.

Both proposals have statistical and regulatory obvious positive consequences, basically by solving the lack of controlled data, and allowing a proper estimation of the treatment effects of the product in both efficacy and safety, thus facilitating interpretation of the study results. The randomized placebo-controlled trial over SOC with sequential analysis for small populations may have positive practical consequences by reducing sample size requirements and facilitating recruitment, with no major ethical drawbacks beyond those related to the use of placebo on top of SOC. Similarly, an RCT trial with sample size reassessment and hypothesis testing in adaptive survival studies might optimise the final sample size with the same practical advantages, by designing a feasible sample-size and giving the opportunity to expand the sample in case of predicting a need for additional precision in the final analysis as derived from the sample size interim assessment.

6.7.2. Ilaris®

Main uncertainties for Ilaris®, (Cluster B: repeated acute episodes), were related to the optimal dose, potential overestimation of the effect and lack of demonstrated benefits beyond short-term control of inflammation.

Specific dose-finding studies were argued as not possible due to the scarcity of patients. Thus, modelling and simulation was applied to data obtained in an open-label dose-titration study in 34 CAPS patients that included assessment of drug levels and disease relapse.
The single pivotal CAPS study (121) used a controlled withdrawal design since it was argued that given the severity of the patient population in this trial and the expectations of efficacy from uncontrolled data (122–124), a placebo-controlled randomization of canakinumab-naïve patients would not be ethical, especially in children.

At the moment of the regulatory assessment, efficacy was demonstrated by the prevention of relapses in MWS patients at short term, and it was expected that this might impact in a better long-term prognosis (reduction/prevention of amyloidosis and end organ damage caused by the inflammatory process). However, and despite the rational appears to be reasonable and sound, this cannot be ascertained unless very long-term trials are conducted, which was deemed unfeasible in this setting.

With regards to the treatment schedules strategies, it cannot be ruled out that other (non-tested) strategies would lead to a better benefit/risk ratio. Nevertheless, given the low patient availability, the severity of the disease and the sound efficacy results, this may be overruled in this setting.

A better estimation of the effect could be addressed by the two methods proposed, i.e. a parallel delayed start pivotal study or a multi-arm group sequential design. Both methods permit to obtain a more understandable and useful prediction of the treatment effect, that due to the exposure to it instead to its withdrawal.

Nevertheless, only the multi-arm group sequential design would be able to address uncertainties on the optimal dose, although this may increase sample size requirements. Uncertainties on the long-term benefits in terms of prevention of amyloidosis and/or end organ damage would remain. It is fair to say that Ilaris® had already used (novel) an alternative and efficient (novel) methodology, making room for improvement smaller.

In essence, no critical advantages would had been expected using the two proposed alternative methods apart from the type of treatment effect estimation (for both new proposed designs) and the possibility to clearly better dose finding information so that theoretically would had minimised exposure to ineffective dose strategies (for the multi-arm group sequential design).

6.7.3. Revestive®

For Revestive®, Cluster C chronic stable/slow progression conditions, a multi-arm group sequential design with a simultaneous stopping rule using a continuous variable, analysed by MMRM adjusted by baseline value and with sample size reassessment, including GAS as a secondary endpoint, will provide a more robust and informative information on clinical relevance than the development plan that was actually done, thus providing better information for benefit/risk decision.

The confirmatory clinical program included two double-blind, randomized, controlled, parallel group trials of 24-week duration, using surrogate endpoints by measuring the needs of parenteral nutrition. The first trial, which was negative, included two active doses and placebo, and the second trial included the low dose from the previous trial and placebo, with statistically significant results (125).

The intermediate variable, amount of parenteral nutrition, was deemed reasonably representative of the clinical goal, i.e. the rate of patients that could be weaned off PN/i.v. fluid completely. Volume requirements, highly dependent on the baseline values, are considered as a relevant source of variability that should ideally be taken into account in the analysis. The dose selection was based on a very small exploratory trial with only 16 patients and was inconclusive due to heterogeneity and lack of standardised protocols for food intake.

Despite the statistical significance in the intermediate variable, it is difficult to foresee in which magnitude the observed benefit can be translated into the clinical goal. In addition, it was not possible to demonstrate any significant difference in QoL. No robust long-term information was available at the time of the regulatory assessment, although there was an on-going trial addressing this point.

Also lack of long-term data poses a safety risk which may include malignancy considering the mechanism of action of teduglutide. A clinical development with a perspective of 1 year (or less) is expected to be unable to detect such risks. However, considering the serious and disabling nature of the condition with a considerable impact on QoL and only limited symptomatic treatment options, the demonstrated effect of the drug was considered to clearly outweigh the safety concerns, whenever treatment is not continued log-term.

The alternative proposed design would adequately address the concerns regarding the handling of heterogeneity of volume requirements, by including that variable as a stratifying factor in the randomisation and as a covariate in the analysis. This would impact in a reduction of variability and a higher statistical power. Also, by using a continuous variable (volume requirements) and the MMRM approach, which would include repeated measurements for the patients, the statistical power would be further optimised.

The uncertainties on the dose selection would be addressed by a multi-arm trial with different dose levels, which might be stopped simultaneously for efficacy or dropped individually for futility. Overall, with one well-designed single-pivotal trial the sample size would not be increased, and the uncertainties regarding the dose selection would be ameliorated. Finally, the assessment of the clinical relevance would be complemented, once efficacy in the PEP shown, by the use of the GAS.

In summary, the proposed alternative method allows reducing the risk of failure due to uncertainties in dose selection based on few patients and increases the sensitivity to changes because of using a continuous variable. The uncertainties about clinical relevance due to the use of an intermediate endpoint and the long-term safety concerns cannot be addressed and would require post-marketing monitoring.

6.7.4. Soliris®

For Soliris®, Cluster D chronic progressive led by one organ/system, a prospectively defined meta-analysis of small trials using two techniques developed within the ASTERIX project for the primary analysis: “Prior distributions for variance parameters in sparse-event meta-analysis” and “Heterogeneity estimators in zero cells meta-analysis” is proposed.

The demonstration of efficacy of Soliris in PNH patients with haemolysis was assessed in a stratified by number of units transfused randomized, double-blind, placebo-controlled 26 weeks pivotal trial (126), using co-primary endpoints (haemoglobin stabilization and units of PRBCs transfused). Supportive evidence came from a single arm 52-week study and from a long-term extension study.

At the time of the design of the pivotal trial, there was only previous evidence from a 12-week, open-label clinical study (n=11), suggesting a reduced intravascular haemolysis and transfusion requirements, and the intermediate variables were considered sound enough to be confirmed in the pivotal trial.

The main limitations/uncertainties in this case are related to the use of short-term surrogate endpoints, with some supportive clinical outcome data on fatigue and QoL. The assessment of efficacy in the prevention of thromboembolic events, which is the main cause of mortality in PNH, is based only on non-controlled clinical trials. Also, no formal dose-finding studies were conducted in the current indication. Finally, there were also uncertainties related to the characterization of the safety profile of Soliris.

Applying a multi-arm multi-stage design with simultaneous stopping rules may provide some practical and ethical advantages (i.e. reducing sample size, time to completion, facilitating enrolment). Also, this design would help to address the concerns regarding the lack of dose selection data extrapolated from other diseases.

As a complementary key strategy, a prospectively defined meta-analysis of small trials based on “Prior distributions for variance parameters in sparse-event meta-analysis” and “Heterogeneity estimators in zero cells meta-analyses” would generate more robust information on the actual effect on of treatment on reducing the incidence of thromboembolic events, which are the critical and main cause of mortality associated to PNH and thus a key relevant clinical objective.

Also, meta-analytic techniques may be applied to the analysis of safety information, considering not only the PNH, but also other indications. Bayesian approaches could be a suitable option for better integration of data, increasing information and improving interpretation at the end of any study, especially if information on the condition is already existing and similar to the one obtained in the trial. This would allow focusing on relevant outcomes, i.e. thromboembolic events, and overall efficacy/safety assessment, in addition to pharmacodynamic markers, which may increase robustness and relevance of the development, by addressing one of the identified drawbacks of the actual development.

Use of meta-analytic techniques would unlikely have major ethical or practical impact at the time of the marketing authorization application, but the development would have much higher

level of evidence for hard endpoints, thus improving the robustness of regulatory decision and reducing risks for the future patients.

6.7.5. Fabrazyme®

For Fabrazyme®, Cluster E: chronic progressive led by multiple organs/systems, an enriched design with more strict inclusion and exclusion criteria, and the use of multiple co-primaries with application of a Fallback test for co-primary endpoints is proposed.

The actual pivotal evidence consisted of a single-pivotal trial strategy. The study was a phase III, double blind, placebo controlled multi-centre study, conducted in 58 patients, followed up to 20 weeks, assessing concentrations of a biomarker, GL-3, in plasma and urine, and complemented with histological examinations of pre- and post-treatment biopsy samples of several organs (127,128). Data from phase I and II clinical trials suggested that enzyme replacement therapy (ERT) with recombinant human enzyme (Fabrazyme®) might reduce lysosomal GL3 deposition (127,129) and the pivotal study was addressed to show activity using pharmacodynamics markers (reduction of sphingolipids in the target organs) with a complete absence of clinical endpoints, i.e. symptoms, function, etc.

Although on a theoretical basis the used biomarkers may precede clinical improvement or a stabilization of the clinical condition, none of the clinical parameters investigated as secondary end-points did reach statistically significant improvement. Thus, the treatment was assessed only at short-term by means of biomarkers, and assuming to provide a long term clinical benefit. In fact, at the time of decision making for marketing authorisation, there were uncertainties on how changes in sphingolipids may later translate into a clinically relevant effect, and inference on the potential benefit of the product was assumed to derive from the hypothesis and physiopathology. Actually, adverse events (rigors, fever and skeletal pain) were more frequent with active treatment than with placebo, so that in clinical terms and according to available data, the effect could even be deleterious. Such uncertainties were acknowledged, so that the authorisation was issued under exceptional circumstances and a number of post-marketing commitments were requested in order to collect additional long- term efficacy and safety data.

The alternative design proposed considers an enriched trial with more strict inclusion and exclusion criteria, aimed to select more severe patients who have more chances to be able to show treatment-related changes vs. placebo, and the use of multiple co-primaries that may be each independently able to conclude confirmation of efficacy by application of a Fallback test for co-primary endpoints.

From other information on the disease and trials with similar treatments, it can be derived that patients with more advanced disease may be more responsive to treatment, so that clinical changes may be quantified better in an enriched population of patients with advanced disease. The advantages of this method are basically that the confirmatory efficiency is increased, as a result of both the likely increase of the effect size by selection of patients, and the reduction of risk of failing on the choice of the primary variable, by using a number of related variables with similar and complementary clinical relevance at the same level of confirmatory validity. Clinical consistency is also likely increased, thus potentially reducing uncertainties at the time of assessment of clinical benefit.

Since Fabry disease is a genetic disorder, the replacement therapy is foreseen to be a life-long therapy, and thus safety is key in this indication. The effect of the alternative design on the robustness of safety data is slightly favorable, since the population is anticipated to be more susceptible to adverse reactions and is the target population where there are more causes for concern about risks from a clinical perspective. The long-term safety should be addressed via post-marketing studies.

Disadvantages may come from the ethical point of view, since the selection of subjects for an enriched design limits the number of subjects who may access the treatment within an experimental setting, and potential delays due to increased recruitment difficulties may lead to longer time to complete the pivotal evidence, and thus later regulatory access to the new drug. As regards to the exposure to placebo, the method is neutral, although less patients may be eligible to participate in the trial because of strict inclusion criteria, as already explained. Considering that the condition has quite a slow progression, this may not seem a critical point with severe prognostic impact.

6.7.6. Opsumit®

For Opsumit® (macitentan) Cluster F staged conditions, the alternative design consists of a development plan including a single-pivotal phase III, multicenter, double-blind, randomised, placebo-controlled, parallel group, event-driven, trial in patients with Pulmonary Arterial Hypertension (PAH). The main endpoint is a composite variable considering events of different clinical relevance, even if all reflect a negative clinical outcome. Individual components would be analysed as secondary endpoints hierarchically and with multiplicity adjustments.

The authorisation is supported by a single long term pivotal phase III with identical characteristics to the alternative proposal except for the PEP. Given the clinical presentation with multiple symptoms, the PEP was a composite variable focused to cover all of them, including also death, hospitalisation, surgery, lung transplantation among a number of components.

Main uncertainties are related to the lack of a formal proof of concept study, of a specific dose-finding data in PAH and to the main endpoint(130). There was availability of appropriate biomarkers to have conducted a pharmacodynamic trial. Binding of an endothelin receptor antagonist to endothelin receptors causes an increase in plasma endothelin-1 levels, which can be used as a marker of pharmacological effect and potency on the endothelin receptor. This effect of ERAs is of rapid onset and thus is a pharmacodynamic measurement of the activity.

No dedicated dose-finding study was conducted in patients with PAH and the Applicant‘s strategy was to employ pharmacodynamic data on plasma endothelin-1 levels and hemodynamic efficacy data on blood pressure reduction in patients with mild to moderate essential hypertension for extrapolation, and thus to determine the doses to be tested in the Phase 3 clinical outcome study in patients with PAH. The underlying assumption was that a dose shown to be efficacious in systemic hypertension would also be hemodynamically effective in PAH, as previously observed with bosentan.

The main endpoint of the pivotal trial, a composite variable considering multiple components of different clinical relevance, was statistically significant. Current guidelines suggested at the time of the design that the primary end point in phase 3 trials of new treatments for pulmonary arterial hypertension should be morbidity and mortality (131–133). However, while differences against placebo were already detected after 6 months of follow-up, longer follow-up was not able to detect effects on more finalist variables such as mortality. Thus, it can be questioned the convenience of combining in the same composite endpoint morbidity and mortality.

With regards to the alternative design for the pivotal trial, given the clinical presentation affecting the use of multiple endpoint analyses, it is suggested to use several co-primary endpoints to be analysed individually implementing Fallback-tests (multiplicity adjustment which increases the chances of finding at least one significant end-point). This is particularly useful when they have different clinically relevance, so increasing the robustness of the evidence presented. Stratification and the use of sequential methods and adaptive approaches are useful ways to increase efficiency in this development.

The original overall 1% two-sided alpha level (applied because only one-single pivotal study was submitted to support the application for marketing authorization) might possibly be relaxed to the standard 5%. A prospectively defined meta-analysis with pre- plus post-MAA studies would have been probably helpful to alleviate the concerns of a single-pivotal study, and thus the sample size might be optimised.

6.7.7. Additional considerations on the simulations exercise

Results of the performed simulations show that (novel) methods applied to the design and analysis of clinical studies of the development plan can be of help in addressing important regulatory uncertainties, which is regarded as one of the major limitations of orphan drug applications. Therefore, (novel) methods and design analyses are considered useful tools to address challenges in orphan drugs developments and increase the ability of generating sufficiently robust evidence that facilitates the regulatory decision-making process. Results also show that the impact on other relevant aspects varies depending on the methods applied, but when negatively impacted overall this is not done in a relevant or unacceptable way.

The results of the simulations drew additional interesting results. In particular, conclusions on applicability and added value of novel methods were extended when changes were not limited to the actual settings of the study design of the pivotal studies, i.e. considering the characteristics of the medical condition and optimising the drug development program, rather than just improving the pivotal trial as presented in isolation (e.g. Defitelio®, Fabrazyme®).

Therefore, novel methods are useful tools not only for the optimization of pivotal trials but rather to optimize the strategical development plan. Thus, to take full advantage of the added value of novel method(s) these should be considered early and for the clinical development program in its totality. In this way, the number of options that can be applied for a given case is broadened, and carefully considering the pros and cons of the different options might facilitate selecting the most suitable method(s) to address main challenges of a particular case. As a consequence, the likelihood to conduct an efficient clinical development able to generate

sufficiently robust evidence that facilitates the regulatory decision-making process might be increased.

In addition, the results of the simulations also showed that for some (novel) methods or designs, in spite of being applicable, the overall added value over the currently conducted one was not so relevant, particularly when some of the major remaining uncertainties could not be addressed (e.g. long-term benefits of Ilaris® in reducing amyloidosis or end organ damage) or when the impact of addressing uncertainties is at the expense of an increase in sample size requirements, which may be difficult for some conditions (e.g. Defitelio® in the treatment of VOD).
Further, in some other cases the small advantages in one aspect were counterbalanced by a negative effect in other relevant aspect, so making the overall added value of applying a particular method of little relevance, e.g. a reduction in sample size requirements might be at the cost of reducing an already small safety database or at the expenses of reducing the chance to see an effect on hard endpoints (e.g. Soliris® in the treatment of PNH).

In summary, recommendations on applicability of methods by clusters of conditions are a useful practical tool when planning the overall clinical development program of a particular OMP. Multiple options may be applicable and have added value on particular aspects of the development. The final decision on the best possible option should consider the particular characteristics of the actual condition under investigation, with the ultimate goal to generate a sufficiently solid, robust and clinically relevant evidence in the most efficient way to substantiate outstanding benefit-risk balance decisions.

6.8. Translation of new methods and applicability into guidance

Limited or null research experience in most OD, and the absence of specific regulatory guidelines, causes uncertainty for drug developers on what to provide with the approval package. Regulators are committed to support developers on the appropriate studies to be conducted in the development of a medicine in order to generate adequate data for the benefit-risk assessment and scientific guidelines are considered a suitable tool. Clinical scientific guidelines are useful EU documents that reflect a harmonized position on the technical and scientific requirements, as agreed by EU Member States, for the demonstration of efficacy and safety of a medicinal product.

Scientific guidelines are aimed to help sponsors in the preparation of the clinical marketing authorisation but also serve to provide advice to competent authorities and/or other interested parties.

The growing role of academia in drug discovery in many areas (134,135), including prominently OD where the early phase of discovery is usually resourced by charitable or traditional modes of academic funding, makes the need for regulatory assistance in the form of publicly available scientific guidelines particularly relevant, in order to guide clinical researchers through the relevant methodological and regulatory requirements.

However, giving specific recommendations to each of the about 6,000 existing rare diseases is not feasible and so, early after the start of the ASTERIX project alterative options were

considered. Grouping of conditions based on e.g. the physiopathology characteristics or the course of the disease, characteristics of methods, etc. had been previously tested and proved to be of limited value. Therefore, the present work has been constructed based on the proposed grouping of medical conditions into 6 clusters defined by characteristics of the disease and the intended therapeutic indication, a broader approach closer to the clinical way of thinking when planning a clinical research (114,136).

Based on the simulation analyses conducted and taking into account the representativeness of the condition within the cluster, drafting general recommendations for each cluster of medical conditions could be done from a more solid ground, where the focus is placed on the complete development plan at an early stage and the added value of novel methodologies could be assessed in broad terms. This has proven to increase the general applicability of the methods, opening the possibility to conduct a broad range of randomized controlled studies, which might result unfeasible in late stages of the development program if the clinical equipoise is deemed lost or simply due to researchers´ and patients´ expectations based on preliminary but likely overestimated promising results from a single arm pivotal trial (20). As pointed out by Lasch F. at al., particularly in rare disease there is merit in planning the research strategy based on RCTs as combinable building blocks that provide unbiased estimates of the treatment effect and, more importantly, avoid undocumented selection of patients. While from a practical and narrow point of view it is tentative to think that single-arm trials are the better choice in diseases with limited patient recruitment, the risk that planning a single-arm trial may generate wasteful information that is, at best, difficult to be used in future research, should be seen as a disincentive for single-arm trials.

Recommendations are intended to guide investigators throughout the different options of (novel) methods that could be applied to the planning and design of a clinical development program for conditions belonging to any of the 6 proposed clusters of medical conditions, incentivizing the most efficient possible use of the limited human recourses to generate a robust package of evidence.

The access to this practical guidance procedure begins when the condition under investigation is assigned to one of the six Clusters of medical conditions. The relevant general characteristics of the condition within the Cluster added to the individual particularities like e.g. the rarity of the condition, the existence of treatment options available, the existence of prior useful information that could be integrated, etc., would allow tailoring the different methodological options for consideration in a particular case.

These are general recommendations aimed to assists sponsors and investigators when planning the development program of a particular OMP. It is not intended to give a single solution to each problem, but rather to guide stakeholders throughout the different suitable options to address the methodological challenges usually faced during the clinical development of an OMP. To this aim, (novel) methods, including those developed within the ASTERIX Project, are deemed valuable tools as long as, if correctly applied, they can provide a good balance of robustness and efficiency in the generated evidence. Therefore, the awareness and use of novel methods should be fostered among stakeholders by improving actual scientific guidelines for the development of medicinal products for OD.

6.9. Strengthens and limitations

Most notable strengths of this research are the fact that final recommendations on the applicability of methods rely on the simulations and not on the initial applicability exercise, as has been done in the past and proven to have limitations. The value of the applicability exercise throughout this work is just limited to the demonstration of the general applicability of the novel methods developed within the ASTERIX Project across the different clusters of conditions. However, in the simulation exercise we went beyond the actual development and designed alternative clinical developments, focusing on clinical relevant outcomes and applying novel methods, designs and analysis techniques, including but not limited to those developed within the ASTERIX Project. We confirmed that the novel methods are applicable to real life developments, and that they have the potential to improve clinical drug development for small populations, directly addressing some of the issues flagged in the 'Guideline for Clinical Trials in Small Populations'.

The novel methods applied to the clinical development program demonstrated to have advantages from the regulatory and statistical point of view, as long as more robust evidence could be generated and relevant uncertainties in prior developments is addressed. But importantly, in some cases this was at the expenses of some practical and ethical disadvantages, e.g. increased number of patients to recruit, more patients exposed to placebo, etc., usually due to applying novel methods that rely on randomized controlled studies. Nevertheless, in most cases these ethical and/or practical disadvantages would be less so or not problems at all, if the novel methods were applied early in development instead of late to just amend the pivotal trial. This reinforces the importance of considering the applicability of methods early in the development, considering the overall clinical plan in its totality.

Our framework represents an approach to practical and structured thought on the planning of clinical trials and analyses that will form the basis of a clinical development plan for medicinal products for OD. Recommendations are intended to guide investigators throughout the different options of (novel) methods that could be applied to the planning and design of a clinical development program for conditions belonging to any of the 6 proposed clusters of medical conditions, promoting the most efficient possible use of the limited human recourses but ensuring generation of robust evidence. To this aim, recommendations are aimed to incentivize the use of randomized controlled studies in any form, as early as possible, and all throughout the clinical development program, taking advantage of the added value that novel methodologies can offer. These are a way to minimize or avoid generating wasteful information by single-arm trials that could prevent the conduct of future methodologically sound research. Recommendations also include advice on the use of meta-analytical or Bayesian approaches, which should also be considered early in the developing, but can help to address particularly difficult situations.

This evaluation also has some limitations. Firstly, due to feasibility reasons only 4-6 EPARs were evaluated for applicability within each cluster. Although it was aimed to select a representative sample of different development approaches within each cluster, the applicability

within these EPARs might not be fully generalisable to all conditions and drug development plans within the cluster, moreover when the representativeness of the examples within the clusters varied substantially. However, the exercise showed that a systematic approach including the definition of the applicability pre-requisites, together with the definition of the general characteristics of the medical conditions included in a given cluster, allows guidance to investigators on whether they could consider a given method or not for a certain type of medical conditions. Based on this exercise, it could be concluded that in each individual case the method's pre-requisites, advantages, and disadvantages should be thoroughly evaluated for adequacy in the full context of the drug development program.

While the exercise of applicability may help to define the best toolbox to consider for a given clinical situation, the implications of the methods may differ between conditions and trials, and it should be judged on a case-by-case basis which one of them is optimal. However, this type of approaches that put the focus on the methods applicable to a given pivotal study have limitations and limited practical utility, given that it is the condition in its broadest term what normally guides and determines the planning of a clinical development, and what should be considered prospectively in its totality.

A further limitation is that the level of detail reported regarding information needed to determine applicability and added value was often limited (e.g. recruitment rates, study timelines, etc.), making it difficult to make a thorough and fully informed judgment on the (in) applicability of the method and their added value, because this depended on the judgment regarding what changes were deemed feasible or not. However, it is plausible to anticipate which would be the recruitment times when designing a plan, and during the analysis of trial feasibility.

Additionally, only OMP with positive opinions were included for both the applicability and the simulation exercises, given the lack of accessible information on the negative opinions. The impossibility to include negative opinions could have influenced the applicability of the methods. However, it was conjectured that in negative opinions there is probably even more potential for improvement.

This work was limited to the European regulatory region. It could have included the assessment of other orphan drugs approved in other regions, notably in the US and Japan for instance, in order to cover more orphan conditions. However, several factors would hamper this approach, mainly the use of different criteria for designation of OMP in the US and in EU (i.e. different prevalence cut-off and including medical devices). Furthermore, detailed data on Japanese clinical developments for OMP were not easily available (61).

Besides ethical or clinical considerations on acceptability of different approaches to trial design may vary across regions and the EU perspective of the research team may not represent nor be appropriate in a different setting. Nevertheless, given the international dimension of orphan drug developments recommendations to increase awareness and foster the use of novel methods to the design and analysis of clinical development programs could be generalized, and ideally applied to any regulatory setting.

Not all challenges reported in EPARs or encountered in trials in rare diseases were covered by the novel methods developed within ASTERIX. One possible avenue for extending this validation exercise based on studies reported in EPARs would be to add on the novel methods developed in the ASTERIX project other study designs and methods applicable to rare diseases available in the literature, as the results here demonstrated that this methods validation exercise works and has potential to be extended.

Further research into methods to address these challenges is needed to improve and optimise drug development to ultimately be able to efficiently develop efficacious and safe treatments for all patients suffering from a rare disease.

6.10. Summary and future prospects

While traditional clinical drug developments targeting prevalent conditions can take the privilege of conducting one clinical trial to answer every relevant question for the demonstration of the drug benefits and risks, the development of most OMP cannot be conceived in the same way. The ultimate goal should always be generating a sufficiently solid, robust and clinically relevant evidence to substantiate benefit-risk balance decisions with the lowest level of uncertainty possible, while making the most efficient use of limited resources available.

Due to patients constrains, clinical trials performed in OD are intended to answer as many as possible relevant questions on the benefits and risks of the medicinal product under investigation, and this justifies the use of novel designs and methods of analysis that address multiplicity. Planning the development program for conditions subject to relevant sample size restrictions, often progressive/seriously debilitating and/or life-threatening disorders, like are most orphan conditions, merits a careful planning of the research strategy to ensure generation of useful and robust evidence on the treatment effect in the most efficient way.

As rightly highlighted by Lasch F. at al., while from a practical and narrow point of view it is tentative to think that single-arm trials are the better choice in diseases with limited patient recruitment, the risk that planning a single-arm trial may generate wasteful information that is, at best, difficult to be used in future research is too high and should be discouraged (20). Thus, all stakeholders involve in clinical research of potentially new treatments, from clinicians that have the first hypothesis to researchers, ethics committees, bodies assessing and deciding research grants, and regulators, we all have the responsibility of generating information that is useful and does not compromise the future development of potentially useful therapies by performing poor designed studies. Lack of resources should not be in general a reason to avoid randomized testing even in very early stages.

Further, as pointed out by Parmar MKB at al., forging ahead with a design that does not offer a proper opportunity to improve outcomes is failing the patients it aims to help, and the price is paid by the public and patients rather than the researchers (111).

Novel methodologies applied to the design of clinical studies in orphan conditions have shown having added value in generating a solid and robust evidence and may address important uncertainties commonly faced during the benefit-risk balance evaluation. Thus, novel

methodologies are useful tools at facilitating the regulatory decision-making process and, subsequently, clinical practice decisions.

The main advantages can be expected when novel methods are considered early in the development, allowing the conduct of randomised controlled trials is most cases and thus the best option to provide unbiased estimates of the treatment effect. The focus should be placed on the complete development program and how novel methodologies, including meta-analytical and Bayesian approaches, can be best applied to ensure that the relevant questions are soundly answered.

Therefore, there is a need to increase awareness and the use of novel methodologies in the planning of orphan drug developments as a way to potentially improve clinical drug development for small populations and directly address some of the issues flagged in the 'Guideline for Clinical Trials in Small Populations'. Given the global development scenario of orphan drugs, the growing role of third parties like clinical researchers in early orphan drug development and the reality of the early access tools, make it needed to strengthen the use of novel methodologies early in development by all stakeholders. To this aim, EU scientific guidelines are considered the most suitable place given its broad accessibility.

Our framework, methods and recommendations represent an approach to practical and structured thought on the planning of clinical trials and analyses that might form the basis of a clinical development plan for OMP and could support sponsors on the appropriate studies to generate robust evidence in the scenario of small populations.

7. CONCLUSIONS

1. Orphan medicinal products face a high-level of uncertainties during drug development and at the time of the regulatory decision-making. There is a huge need to strengthen support to developers on the appropriate studies to be conducted in the development of orphan medicine products in order to generate adequate evidence for the demonstration of the benefits and risks.

2. Existing EU scientific guidelines offer limited, fragmented, and not always consistent guidance to provide references on the design of studies for the demonstration of efficacy (and safety) of medicinal products in small populations suffering from a rare disease. Specific guidance is needed on controversial aspects in order to reduce the methodological flaws often encountered in orphan drug developments.

3. Novel methods developed in ASTERIX, including methods for trial design, analysis or meta-analysis of trials in small populations, have an overall limited applicability when the focus is placed on improving the pivotal studies of the actually conducted development programs, keeping the primary development objectives intact.

4. Applicability of novel methods, including those developed within the ASTERIX Project, is extended and their added value in generating a solid and robust evidence is increased, when methods are prospectively applied to the design and analysis of the overall clinical development plan, and can be useful to address important regulatory uncertainties commonly faced during the benefit-risk balance evaluation. Thus, novel methodologies are useful tools at facilitating the regulatory decision-making process and, subsequently, clinical practice decisions.

5. To take full advantage of the added value of novel method(s) these should be considered early in the process of planning the clinical development program. In this way, the number of options that can be applied for a given case is broadened, including the options for randomized controlled studies.

6. Novel methods, including those developed within the ASTERIX Project, are deemed valuable tools as long as, if correctly applied, they can provide a good balance of robustness and efficiency in the generated evidence. Therefore, the awareness and use of novel methods should be fostered among stakeholders, including clinical researchers given the growing role of academia in early drug development in OD.

7. To strengthen the use of novel methodologies early in development and assist drug developers to select the optimal methods for the evaluation of the condition being targeted, it is critical that a revision of the relevant EU scientific guidelines is done, given its broad accessibility to all stakeholders involved in orphan drug research.

8. Planning the development program for conditions subject to relevant sample size restrictions, often progressive/seriously debilitating and/or life-threatening disorders, like are most orphan conditions, merits a careful planning of the research strategy to ensure generation of useful and robust evidence on the treatment effect in the most efficient way even from the very initial exploratory studies.

9. Grouping medical conditions into 6 clusters defined by characteristics of the disease and the intended therapeutic indication, as proposed within the ASTERIX Project, is a practical tool to provide guidance throughout the different methods and design analysis and a suitable one as long as it matches the regulator´s and academia researcher´s point of view, since giving specific recommendations to each of the about 6,000 existing rare diseases is not feasible.

10. Our framework, methods and recommendations represent an approach to practical and structured thought on the planning of clinical developments for products aimed to treat small populations and could support sponsors on the selection of the most appropriate studies to generate robust evidence in the field of OMP.

11. All stakeholders involved in clinical research of potentially new treatments, from clinicians that have the first hypothesis to researchers, ethics committees, bodies assessing and deciding research grants, and regulators, share the responsibility of generating information that is useful and does not compromise the future development of potentially useful therapies by performing poorly designed studies.

8. REFERENCES

1. Orphan medicines [Internet]. [cited 2018 Jun 25]. Available from: https://ec.europa.eu/health/human-use/orphan-medicines_en
2. Orphanet, The portal for rare diseases and orphan drugs [Internet]. [cited 2018 Jun 25]. Available from: https://www.orpha.net/consor/cgi-bin/Education_AboutRareDiseases.php?lng=EN
3. Franco P. Orphan drugs: the regulatory environment. Drug Discov Today. 2013;18(3–4):163–72.
4. EMA Orphan designation [Internet]. [cited 2018 Jun 25]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000029.jsp&mid=WC0b01ac0580b18a41
5. EURORDIS, Rare diseases Europe [Internet]. [cited 2018 Jun 25]. Available from: https://www.eurordis.org/content/what-rare-disease
6. Orphanet, the list of rare diseases [Internet]. [cited 2018 Jun 25]. Available from: https://www.orpha.net/consor/cgi-bin/Drugs_ListOrphanDrugs.php?lng=EN
7. Murphy M. Rare Diseases: Meeting the Unique Challenges of Orphan Drug Development. Applied Clinical Trials [Internet]. 2016;18–20. Available from: http://www.appliedclinicaltrialsonline.com/rare-diseases-meeting-unique-challenges-orphan-drug-development
8. Janet Woodcock MD. FDA, the more we know about rare diseases... [Internet]. [cited 2018 Jun 25]. Available from: https://blogs.fda.gov/fdavoice/index.php/2014/10/the-more-we-know-about-rare-diseases-the-more-likely-we-are-to-find-safe-and-effective-treatments/
9. European Reference Network. European Commission. Luxembourg: Publications Office of the European Union, 2017 [Internet]. Available from: https://ec.europa.eu/health/sites/health/files/ern/docs/2017_brochure_en.pdf
10. Rath A, Salamon V, Peixoto S, Hivert V, Laville M, Segrestin B, et al. A systematic literature review of evidence- based clinical practice for rare diseases : what are the perceived and real barriers for improving the evidence and how can they be overcome ? 2017;1–24.
11. EURORDIS. Rare disease patients' participation in research - A Rare Barometer survey. [Internet]. [Internet]. [cited 2018 Jul 7]. Available from: http://download2.eurordis.org.s3.amazonaws.com/rbv/2018_02_12_rdd-research-survey-analysis.pdf
12. EURORDIS. Survey on "Rare Disease Patient Organisations in Research: their role and priorities for the future". [Internet] . [Internet]. [cited 2018 Jul 7]. Available from: https://www.eurordis.org/sites/default/files/publications/3_FBignami_RDD2010.pdf

13. Genetic Alliance UK. Supporting families affected by a syndrome without a name. [Internet]. [cited 2018 Jul 7]. Available from: https://www.undiagnosed.org.uk/news-events/news/introducing-swan-europe/
14. Pocock SJ. Clinical Trials: A Practical Approach. Chichester [West Sussex]: John Wiley & Sons; 1987.
15. Bell SA, Tudur Smith C. A comparison of interventional clinical trials in rare versus non-rare diseases: an analysis of ClinicalTrials.gov. Orphanet J Rare Dis. 2014;9:170.
16. Kesselheim AS, Myers JA, Avorn J. Characteristics of clinical trials to support approval of orphan vs nonorphan drugs for cancer. JAMA - J Am Med Assoc. 2011;305(22):2320–6.
17. Mitsumoto J et al. Pivotal studies of orphan drugs approved for neurological diseases. Ann Neurol [Internet]. 66(2):184–90. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/ana.21676
18. The needs of the few. Nature [Internet]. 2010;466(7303):160. Available from: https://www.nature.com/articles/466160a.pdf%0D%0A
19. Augustine EF, Adams HR, Mink JW. Clinical Trials in Rare Disease: Challenges and Opportunities. J Child Neurol [Internet]. 2013 Sep 1;28(9):1142–50. Available from: https://doi.org/10.1177/0883073813495959
20. Lasch F, Weber K, Chao MM, Koch A. 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 J Rare Dis. 2017;12(1):1–6.
21. Franklin G. Miller, Ph.D., and Steven Joffe, M.D. MP. Equipoise and the Dilemma of Randomized Clinical Trials. N Engl J Med [Internet]. 2011;364:476–80. Available from: https://www.nejm.org/doi/full/10.1056/NEJMsb1011301
22. International Conference on Harmonization. ICH Topic E 1: Population Exposure: The Extent of Population Exposure to Assess Clinical Safety. European Medicines Agency; CPMP/ICH/375/95. [Internet] . [Internet]. 1995 [cited 2018 Jul 7]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002747.pdf
23. The European Parliament and the Council of the European Union. Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products. Off J Eur Union [Internet]. 2000;L 18:1–5. Available from: https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2000_141_cons-2009-07/reg_2000_141_cons-2009-07_en.pdf
24. Regnstrom J, Koenig F, Aronsson B, Reimer T, Svendsen K, Tsigkos S, et al. Factors associated with success of market authorisation applications for pharmaceutical drugs

submitted to the European Medicines Agency. Eur J Clin Pharmacol [Internet]. 2009;66(1):39. Available from: https://doi.org/10.1007/s00228-009-0756-y

25. The European Parliament and the Council of the European Union, Parliament THEE, Council THE, The OF, Union E. Regulation (EC) No 726/2004 of the European Parliament and of the Council. Off J Eur Union [Internet]. 2004;136(1):1–33. Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32004R0726:EN:HTML
26. European Medicines Agency. Scientific advice and Protocol assistance [Internet]. [cited 2018 Jun 26]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000049.jsp&mid=WC0b01ac05800229b9
27. Tankovic G, Reljanović M, Tankovic G. Designing robust clinical trials for orphan drugs. In: Orphan Drugs [Internet]. Elsevier; 2013 [cited 2018 Jun 25]. p. 187–246. Available from: http://linkinghub.elsevier.com/retrieve/pii/B9781907568091500069
28. Farmaindustria. Cuánto cuesta desarrollar un medicamento [Internet]. 2012 [cited 2018 Jun 25]. Available from: http://www.farmaindustria.es/web/infografia/cuanto-cuesta-desarrollar-un-medicamento/
29. Karnath L. Patent-to-launch time for orphan drugs is 2.3 years longer vs. other drugs [Internet]. 2018. Available from: https://www.apnews.com/60220de1c3e7d0a0f8cad2021111f849
30. Michel M TM. Access to orphan drugs in Europe: current and future issues. Expert Reviews 12(1):23–9. Pharmacoeconomics Outcomes Res. 2012;12(1):23–9.
31. Gammie T, Lu CY, Ud-Din Babar Z. Access to orphan drugs: A comprehensive review of legislations, regulations and policies in 35 countries. PLoS ONE. 2015.
32. Denis A, Mergaert L, Fostier C, Cleemput I, Simoens S. Issues surrounding orphan disease and orphan drug policies in Europe. Appl Health Econ Health Policy [Internet]. 2010;8(5):343–50. Available from: https://doi.org/10.2165/11536990-000000000-00000
33. Houlton S. Orphan medicines: the high cost of hope [Internet]. 2018. Available from: https://www.prescriber.co.uk/article/orphan-medicines-high-cost-hope/
34. Liikanen E. Regulation (EC) 847/2000. Off J Eur Communities [Internet]. 2000;L 103/5(847):5–8. Available from: https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2000_847/reg_2000_847_en.pdf
35. European Parliament, Council of the European Union. Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use. Eudralex [Internet]. 2006 [cited 2018 Jul

7]. p. 6–9. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1530958516699&uri=CELEX:32001L0083

36. Parliament THEE, Council THE, The OF, Union P. 27.12.2006. 2006;
37. European Commission. Commission Regulation (EC) No 2049/2005 of 15 December 2005 laying down, pursuant to Regulation (EC) No 726/2004 of the European Parliament and of the Council, rules regarding the payment of fees to, and the receipt of administrative . Regulation [Internet]. 2005;2005(0):4. Available from: http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32005R2049
38. Lincker, Helene & Ziogas, Constantinos & Carr, Melanie & Porta, Nuria & Eichler H-G. Regulatory watch: Where do new medicines originate from in the EU?. Nat Rev Drug Discov. 2014;13:92–3.
39. European Medicines Agency. Orphan designation: Marketing Authorisation. [Internet] [Internet]. [cited 2018 Jul 7]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_001776.jsp&mid=WC0b01ac0580b18195
40. Hofer MP, Hedman H, Mavris M, Koenig F, Vetter T, Posch M, et al. Marketing authorisation of orphan medicines in Europe from 2000 to 2013. 2018;23(2).
41. EMA, Conditional Marketing authorisation and authorisation under exceptional circunstances [Internet]. [cited 2018 Jun 26]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000925.jsp&mid=WC0b01ac05809f843b
42. Pontes C, Secondary CA, Author C, Pontes C. Orphanet Journal of Rare Diseases Evidence supporting regulatory-decision making on OMP authorisation in Europe : methodological uncertainties. 2013;
43. Picavet E, Cassiman D, Hollak CE, Maertens JA, Simoens S. Clinical evidence for orphan medicinal products- a cause for concern ? 2013;1–8.
44. European Medicines Agency: Product Development Scientific Support Department. European Medicines Agency guidance for applicants seeking scientific advice and protocol assistance. EMA/4260/2001 Rev. 9, 2017. [Internet] [Internet]. [cited 2018 Jul 7]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2009/10/WC500004089.pdf
45. European Medicines Agency. General principles EMA-FDA parallel scientific advice (human medicinal products). EMA/309801/2017 [Internet] [Internet]. [cited 2018 Jul 7]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2009/11/WC500014868

.pdf

46. EUneHTA, European Network for Health Tecnology Assessment [Internet]. [cited 2018 Jun 26]. Available from: https://www.eunethta.eu
47. Tafuri G, Lucas I, Estevão S, Moseley J, D'Andon A, Bruehl H, et al. The impact of parallel regulatory-health technology assessment scientific advice on clinical development. Assessing the uptake of regulatory and health technology assessment recommendations. Br J Clin Pharmacol [Internet]. 2018 May;84(5):1013–9. Available from: http://doi.wiley.com/10.1111/bcp.13524
48. Scientific Guidelines [Internet]. [cited 2018 Jun 26]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000043.jsp&mid=WC0b01ac05800240cb
49. EMA. Guideline on clinical trials in small populations [Internet]. Report No.: Doc. Ref. CHMP/EWP/83561/2005. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf
50. EMA. Annual Report 2017 [Internet]. Annual Report. 2017. Available from: https://www.cb.is/library/Skraarsafn---EN/Annual-Report/Annual_Report_2017.pdf
51. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J, et al. Clinical research for rare disease: Opportunities, challenges, and solutions. Mol Genet Metab [Internet]. 2009 Jan;96(1):20–6. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1096719208002539
52. Gagne JJ, Thompson L, O'Keefe K, Kesselheim AS. Innovative research methods for studying treatments for rare diseases: methodological review. Bmj [Internet]. 2014;349(nov24 15):g6802–g6802. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.g6802
53. Joppi R, Bertele V, Garattini S. Orphan drug development is progressing too slowly. Br J Clin Pharmacol. 2006;61(3):355–60.
54. FP7-HEALTH-2013-INNOVATION-1 framework [Internet]. [cited 2016 Jun 8]. Available from: https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/%0D%0Afp7/calls/fp7-health-2013-innovation-1.html.%0D%0A
55. Roes, Kit and Koch, Armin principal investigators. ASTERIX Project, Advances in Small Trials dEsign for Regulatory Innovation and eXcellence [Internet]. [cited 2018 Jun 25]. Available from: http://www.asterix-fp7.eu
56. Ralf-Dieter Hilgers principal investigator. IDEAL, Integrated Design and Analysis of small population group trials [Internet]. Available from: https://www.ideal.rwth-

aachen.de

57. Nigel Stallard principal investigator. InSPIRe, Innovation in Small Populations Research [Internet]. [cited 2018 Jun 26]. Available from: https://warwick.ac.uk/fac/med/research/hscience/stats/currentprojects/inspire/
58. Ideal Project. Ideal recommendations for small population clinical trials. Orphanet J Rare Dis [Internet]. Available from: https://www.ideal.rwth-aachen.de/?s=Ideal+recommendations+for+small+population+clinical+trials
59. Warwick Medical School. Innovative methodology for small populations research (InSPiRe). 2016; Available from: http://www2.warwick.ac.uk/fac/med/research/hscience/stats/currentprojects/inspire/
60. Hilgers R, Roes K, Stallard N. Directions for new developments on statistical design and analysis of small population group trials. 2016;1–10.
61. Pontes C. Orphanet Journal of Rare Diseases Applicability and added value of novel methods to improve drug development in rare diseases. 2013;
62. Orfali M. Raising Orphans: How Clinical Development Programs of Drugs for Rare and Common Diseases Are Different. CClinical Pharmacol Ther. 2012;92(2):262–264.
63. Richey EA, Lyons EA, Nebeker JR, Shankaran V, McKoy JM, Luu TH, et al. Accelerated Approval of Cancer Drugs: Improved Access to Therapeutic Breakthroughs or Early Release of Unsafe and Ineffective Drugs? J Clin Oncol [Internet]. 2009 Sep 10;27(26):4398–405. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2744277/
64. Roes, Kit, co-chair Koch A. Asterix Project Methodology [Internet]. [cited 2018 Jun 26]. Available from: http://www.asterix-fp7.eu/methodology/
65. Community Research and Development Information System (CORDIS). Advances in Small Trials dEsign for Regulatory Innovation and eXcellence: Final summary report. [Internet]. [cited 2018 Jul 7]. Available from: https://cordis.europa.eu/docs/results/603/603160/final1-final-summary-report-29nov2017.pdf
66. Nair R, Aggarwal R, Khanna D. NIH Public Access. Methods Form Consens Classif Criteria Guidel Dev Semin Arthritis Rheum. 2011;41(2):95–105.
67. Fontanet, JM, Pontes C (Dir) TP (Dir). Study to characterize the designs of clinical trials conducted with orphan medicinal products. Universitat Autònoma de Barcelona (Doctoral Thesis, September 2017. Universitat Autònoma de Barcelona; 2017.
68. FDA. FDA Approved Drug Products [Internet]. [cited 2017 Jan 6]. Available from: https://www.accessdata.fda.gov/scripts/cder/daf/
69. EMA. EMA Clinical data [Internet]. [cited 2017 Jul 17]. Available from:

https://clinicaldata.ema.europa.eu/web/cdp/home

70. FDA. FDA Clinical Trials Register [Internet]. [cited 2017 Oct 16]. Available from: https://www.clinicaltrials.gov/ct2/help/how-find/find-study-results
71. Ristl R, Frommlet F, Koch A, Posch M. Fallback tests for co-primary endpoints. Stat Med. 2016;35(16):2669–86.
72. Magirr D, Jaki T, Koenig F, Posch M. Sample size reassessment and hypothesis testing in adaptive survival trials. PLoS One. 2016;11(2):1–14.
73. Nikolakopoulos S, Roes KCB, van der Tweel I. Sequential designs with small samples: Evaluation and recommendations for normal responses. Stat Methods Med Res. 2018;27(4):1115–27.
74. Nikolakopoulos S, van der Tweel I, Roes KCB. Dynamic borrowing through empirical power priors that control type I error. Biometrics [Internet]. 2017 Dec 11; Available from: http://doi.wiley.com/10.1111/biom.12835
75. Urach S, Gaasterland CMW, Posch M, Jilma B, Roes K, Rosenkranz G, et al. Statistical analysis of Goal Attainment Scaling endpoints in randomised trials. Stat Methods Med Res [Internet]. 2018 Jun 19;0962280218777896. Available from: https://doi.org/10.1177/0962280218777896
76. Ristl R, Xi D, Glimm E, Posch M. Optimal exact tests for multiple binary endpoints. Comput Stat Data Anal [Internet]. 2018 Jun;122:1–17. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0167947318300021
77. Hlavin G, Koenig F, Male C, Posch M, Bauer P. Evidence, eminence and extrapolation. Stat Med. 2016;35(13):2117–32.
78. Brakenhoff TB, Roes KCB, Nikolakopoulos S. Bayesian sample size re-estimation using power priors. Stat Methods Med Res [Internet]. 2018 May 2;0962280218772315. Available from: https://doi.org/10.1177/0962280218772315
79. Urach S, Posch M. Multi-arm group sequential designs with a simultaneous stopping rule. Stat Med. 2016;35(30):5536–50.
80. Pateras K, Nikolakopoulos S, Roes K. Data-generating models of dichotomous outcomes: Heterogeneity in simulation studies for a random-effects meta-analysis. Stat Med [Internet]. 2018 Mar 30;37(7):1115–24. Available from: http://doi.wiley.com/10.1002/sim.7569
81. Spineli LM, Jenz E, Großhennig A, Koch A. 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;12(1):1–7.
82. Gaasterland CMW, Jansen-van der Weide MC, Weinreich SS, van der Lee JH. A systematic review to investigate the measurement properties of goal attainment scaling,

towards use in drug trials. BMC Med Res Methodol [Internet]. 2016;16(1):99. Available from: http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-016-0205-4

83. Pateras K, Nikolakopoulos S, Mavridis D, Roes KCB. Interval estimation of the overall treatment effect in a meta-analysis of a few small studies with zero events. Contemp Clin Trials Commun. 2018;9(October 2017):98–107.
84. EU Clinical Trials Registry [Internet]. [cited 2017 Aug 28]. Available from: https://www.clinicaltrialsregister.eu
85. Spilker B. Guide to Drug Development: A Comprehensive Review and Assessment, 1s. Ed. Lippincott Williams & Wilkins 2008;
86. Pesiou S. Master thesis: Review of current European regulatory guidance relevant to the clinical development of medicinal products aimed to treat rare medical conditions. Universitat Autònoma de Barcelona, Bellaterra (Barcelona) September 2017. 2017.
87. CHMP European Public Assessment Report for Defitelio, INN-defibrotide. EMA [Internet]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002393/WC500153152.pdf
88. CHMP European Public Assessment Report for Ilaris, INN-canakinumab. EMA. 2009;(July):1–43. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/001109/WC500031679.pdf
89. CHMP European Public Assessment Report for Revestive, INN-teduglutide. EMA [Internet]. Vol. 44, European Medicines Agency. 2012. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002345/WC500132928.pdf
90. CHMP European Public Assessment Report for Soliris, INN-eculizumab. EMA [Internet]. 2007. p. 1–41. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Scientific_Discussion/human/000791/WC500054212.pdf
91. CHMP European Public Assessment Report for Fabrazyme, INN-agalsidase beta. EMA [Internet]. Vol. 0, Development. 2004. p. 1–16. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Scientific_Discussion/human/000370/WC500020543.pdf
92. CHMP European Public Assessment Report for Opsumit, INN-macitentan.EMA [Internet]. Vol. 44. 2013. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/0026

97/human_med_001717.jsp&mid=WC0b01ac058001d124

93. CHMP assessment report templates. EMA [Internet]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/document_listing/document_listing_000337.jsp&mid=WC0b01ac0580022719#section5
94. Halpern SD. The Continuing Unethical Conduct of Underpowered Clinical Trials. JAMA [Internet]. 2002 Jul 17;288(3):358. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.288.3.358
95. Niraula S, Seruga B, Ocana A, Shao T, Goldstein R, Tannock IF, et al. The Price We Pay for Progress: A Meta-Analysis of Harms of Newly Approved Anticancer Drugs. J Clin Oncol [Internet]. 2012 Aug 20;30(24):3012–9. Available from: https://doi.org/10.1200/JCO.2011.40.3824
96. Arnardottir AH, Haaijer-Ruskamp FM, Straus SMJ, Eichler HG, de Graeff PA, Mol PGM. Additional safety risk to exceptionally approved drugs in Europe? Br J Clin Pharmacol. 2011;72(3):490–9.
97. Picavet E, Cassiman D, Hollak CE, Maertens JA, Simoens S. Clinical evidence for orphan medicinal products-a cause for concern? Orphanet J Rare Dis [Internet]. 2013 Oct;8(1):164. Available from: https://doi.org/10.1186/1750-1172-8-164
98. Dupont AG, Van Wilder PB. Access to orphan drugs despite poor quality of clinical evidence. Br J Clin Pharmacol [Internet]. 2011 Apr 1;71(4):488–96. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3080635/
99. Adaptive pathways [Internet]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000601.jsp&mid=WC0b01ac05807d58ce
100. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: Analysis of EMA and FDA approvals 1999-2014. BMJ Open. 2016;6(6).
101. Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13. BMJ [Internet]. 2017 Oct 4;359. Available from: http://www.bmj.com/content/359/bmj.j4530.abstract
102. Eichler H-G, Pignatti F, Flamion B, Leufkens H, Breckenridge A. Balancing early market access to new drugs with the need for benefit/risk data: a mounting dilemma. Nat Rev Drug Discov [Internet]. 2008 Sep 12;7:818. Available from: http://dx.doi.org/10.1038/nrd2664
103. Pignatti F, Jonsson B, Blumenthal G, Justice R. Assessment of benefits and risks in

development of targeted therapies for cancer - The view of regulatory authorities. Mol Oncol [Internet]. 2015 May;9(5):1034–41. Available from: http://doi.wiley.com/10.1016/j.molonc.2014.10.003

104. Hofer MP, Jakobsson C, Zafiropoulos N, Vamvakas S, Vetter T, Regnstrom J, et al. Impact of scientific advice from the European Medicines Agency. Nat Rev Drug Discov [Internet]. 2015 May 17;14(5):302–3. Available from: http://www.nature.com/articles/nrd4621
105. Valsecchi M.G. DLP. Strategies for Trial Design and Analyses. In: Saha V., Kearns P. (eds) New Agents for the Treatment of Acute Lymphoblastic Leukemia. In: Springer, New York, NY. 2011. p. 83–104.
106. Bogaerts J, Sydes MR, Keat N, McConnell A, Benson A, Ho A, et al. Clinical trial designs for rare diseases: Studies developed and discussed by the International Rare Cancers Initiative. Eur J Cancer. 2015;51(3):271–81.
107. Huang B, Giannini EH, Lovell DJ, Ding L, Liu Y, Hashkes PJ. Enhancing crossover trial design for rare diseases: Limiting ineffective exposure and increasing study power by enabling patient choice to escape early. Contemp Clin Trials [Internet]. 2014 Jul 1;38(2):204–12. Available from: http://dx.doi.org/10.1016/j.cct.2014.05.001
108. Hampson L V., Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. Stat Med. 2014;33(24):4186–201.
109. Cornu C, Kassai B, Fisch R, Chiron C, Alberti C, Guerrini R, et al. Experimental designs for small randomised clinical trials: an algorithm for choice. Orphanet J Rare Dis [Internet]. 2013 Mar;8(1):48. Available from: https://doi.org/10.1186/1750-1172-8-48
110. Gupta S, Faughnan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare diseases. J Clin Epidemiol [Internet]. 2011 Oct 1;64(10):1085–94. Available from: http://dx.doi.org/10.1016/j.jclinepi.2010.12.019
111. Parmar MKB, Sydes MR, Morris TP. How do you design randomised trials for smaller populations? A framework. BMC Med [Internet]. 2016;14(1):1–10. Available from: http://dx.doi.org/10.1186/s12916-016-0722-3
112. Lusine A, M. FB, George T, E. FM, R. JS, R. DI, et al. Alternative designs for clinical trials in rare diseases. Am J Med Genet Part C Semin Med Genet [Internet]. 2016 Nov 14;172(4):313–31. Available from: https://doi.org/10.1002/ajmg.c.31533
113. Whicher D, Philbin S, Aronson N. An overview of the impact of rare disease characteristics on research methodology. Orphanet J Rare Dis. 2018;13(1):1–12.
114. C. Pontes. Conditions clustering (to appear).
115. Pontes C, Fontanet JM, Gomez-Valent M, Rios Guillermo J, Vives Vilagut R, Morros R,

et al. Clustering of rare medical conditions based on clinical features which determine applicability of investigative designs and methods to their study. Clin Ther [Internet]. 2015 Aug;37(8):e11–2. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0149291815003501

116. Abrahamyan L, Feldman BM, Tomlinson G, Faughnan ME, Johnson SR, Diamond IR, et al. Alternative designs for clinical trials in rare diseases. Am J Med Genet Part C Semin Med Genet. 2016;172(4):313–31.
117. Richardson P, Linden E, Revta C, Ho V. Use of defibrotide in the treatment and prevention of veno-occlusive disease. Expert Rev Hematol [Internet]. 2009 Aug 10;2(4):365–76. Available from: http://www.tandfonline.com/doi/full/10.1586/ehm.09.30
118. Chalmers TC. Randomization of the First Patient. Med Clin North Am [Internet]. 1975 Jul;59(4):1035–8. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0025712516320016
119. Sibbald B, Roland M. Understanding controlled trials: Why are randomised controlled trials important? BMJ [Internet]. 1998 Jan 17;316(7126):201–201. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.316.7126.201
120. Haffner ME, Whitley J, Moses M. Two decades of orphan product development. Nat Rev Drug Discov [Internet]. 2002 Oct 1;1(10):821–5. Available from: http://www.nature.com/articles/nrd919
121. Lachmann HJ, Kone-Paut I, Kuemmerle-Deschner JB, Leslie KS, Hachulla E, Quartier P, et al. Use of Canakinumab in the Cryopyrin-Associated Periodic Syndrome. N Engl J Med [Internet]. 2009 Jun 4;360(23):2416–25. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa0810787
122. Church LD, Savic S, McDermott MF. Long term management of patients with cryopyrin-associated periodic syndromes (CAPS): focus on rilonacept (IL-1 Trap). Biologics [Internet]. 2008 Dec;2(4):733–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19707454
123. Kuemmerle-Deschner, Jasmin; Tzaribachev, Nikolay; Hausmann, S; Gramlich, K; Felix, Sandra; Diaz-Lorente, M; Giesen, A; Rordorf C. Long-lasting response to ACZ885 (a new human IgG1 anti-IL1 beta monoclonal antibody) in patients in Muckle-Wells Syndrome (MWS). In: Clinical and experimental rheumatology 26 (2). p. 180–180.
124. Church LD, McDermott MF. Canakinumab: a human anti-IL-1β monoclonal antibody for the treatment of cryopyrin-associated periodic syndromes. Expert Rev Clin Immunol [Internet]. 2010 Nov 10;6(6):831–41. Available from: http://www.tandfonline.com/doi/full/10.1586/eci.10.66

125. Vipperla K, O'Keefe SJ. Study of teduglutide effectiveness in parenteral nutrition-dependent short-bowel syndrome subjects. Expert Rev Gastroenterol Hepatol [Internet]. 2013 Nov 10;7(8):683–7. Available from: http://www.tandfonline.com/doi/full/10.1586/17474124.2013.842894

126. Hillmen P, Young NS, Schubert J, Brodsky RA, Socié G, Muus P, et al. The Complement Inhibitor Eculizumab in Paroxysmal Nocturnal Hemoglobinuria. N Engl J Med [Internet]. 2006 Sep 21;355(12):1233–43. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa061648

127. Schiffmann R, Murray GJ, Treco D, Daniel P, Sellos-Moura M, Myers M, et al. Infusion of alpha -galactosidase A reduces tissue globotriaosylceramide storage in patients with Fabry disease. Proc Natl Acad Sci [Internet]. 2000 Jan 4;97(1):365–70. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.97.1.365

128. Eng CM, Guffon N, Wilcox WR, Germain DP, Lee P, Waldek S, et al. Safety and Efficacy of Recombinant Human α-Galactosidase A Replacement Therapy in Fabry's Disease. N Engl J Med [Internet]. 2001 Jul 5;345(1):9–16. Available from: http://www.nejm.org/doi/abs/10.1056/NEJM200107053450102

129. Eng CM, Banikazemi M, Gordon RE, Goldman M, Phelps R, Kim L, et al. A Phase 1/2 Clinical Trial of Enzyme Replacement in Fabry Disease: Pharmacokinetic, Substrate Clearance, and Safety Studies. Am J Hum Genet [Internet]. 2001 Mar;68(3):711–22. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0002929707631103

130. Pulido T, Adzerikho I, Channick RN, Delcroix M, Galiè N, Ghofrani H-A, et al. Macitentan and Morbidity and Mortality in Pulmonary Arterial Hypertension. N Engl J Med [Internet]. 2013 Aug 29;369(9):809–18. Available from: http://www.nejm.org/doi/10.1056/NEJMoa1213917

131. McLaughlin V V., Badesch DB, Delcroix M, Fleming TR, Gaine SP, Galiè N, et al. End Points and Clinical Trial Design in Pulmonary Arterial Hypertension. J Am Coll Cardiol [Internet]. 2009 Jun;54(1):S97–107. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0735109709012091

132. Committee for Medicinal Products for Human Use. Guideline on the clinical investigations of medicinal products for the treatment of pulmonary arterial hypertension. London: European Medicines Agency, 2009 ([Internet]. [cited 2018 Jul 8]. Available from: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/12/WC500016686.pdf

133. Galiè N, Simonneau G, Barst RJ, Badesch D, Rubin L. Clinical worsening in trials of pulmonary arterial hypertension: results and implications. Curr Opin Pulm Med

[Internet]. 2010 May;16(Suppl 1):S11–9. Available from: https://insights.ovid.com/crossref?an=00063198-201005001-00002

134. Palmer M, Chaguturu R. Academia–pharma partnerships for novel drug discovery: essential or nice to have? Expert Opin Drug Discov [Internet]. 2017;12(6):537–40. Available from: https://doi.org/10.1080/17460441.2017.1318124

135. Frearson J, Wyatt P. Europe PMC Funders Group Drug Discovery in Academia- the third way ? 2011;5(077705):1–14.

136. ASTERIX_Deliverable WP5_5_1_2015-05-29 [Internet]. Available from: https://cordis.europa.eu/result/rcn/219205_en.html

9. INDEX OF TABLES

10. INDEX OF FIGURES

11. ANNEXES

INDEX OF ANNEXES

Annex 1. List of characteristics used to build the studies profile

#	Characteristic
a.1.1.	Number of arms in main trial(s)
a.1.2.	Interim analysis Y/N, if so, reason: stopping for futility, overwhelming evidence of efficacy, safety.
a.1.3.	Type of endpoint (PEP) (binary, continuous) Composite? Time to event?
a.1.4.	Type of (major) secondary endpoints (fill in as above)
a.1.5.	Adaptive randomisation? Detail
a.1.6.	Delta time= recruitment - assessment (delay) /immediate or delayed response
a.1.7.	Recruitment rate
a.1.8.	Seamless/adaptive design?
a.1.9.	Allocation ratio?
a.1.10.	Did they allow dropping of arms?
a.1.11.	What was the control group? Add-on?
a.1.12.	MRCT? Multicentric? How and how many?
a.1.13.	Summary of models used in planning (e.g., disease progression, dropout, dose–response)
a.2.1	type control and sample size - Detail
a.2.2.	Justification of design (i.e. use or not of control, what control and how?)
a.2.3.	(Blinding of) sample size reassessment?
a.2.4.	Immediate/Delayed responses +/- recruitment rate
a.2.5.	Correction of rejection boundaries (in case of small samples)?
a.3.1.	Sequential trial? With a maximum sample size?
a.3.2.	Disease severity and seriousness
a.3.3.	Available treatment options/Ranking and selection (BSC, SOC, other treatment or strategy?)
a.3.4.	Possible toxicity of the treatment under consideration
a.3.5.	available sample size/Maximum sample size estimation and consideration (i.e. patient horizon)
a.4.1.	Type of endpoint - (continuous, binary)
a.4.2.	Sequential design? Can it be designed as sequential? Interim analyses? How many? Detail
a.4.3.	Randomisation considerations? Randomised? How?
a.5.1.	Control arm? Justification for use/no use
a.5.2.	Available data from previous similar study/ies? (estimated variance from prior (pilot) study?)
a.5.3.	(Ideally) the same sample size per group in old and new trial?
b.1.1.	Slowly and constantly progressive disease?
b.1.2.	Placebo comparator?
b.1.3.	Intervention has lasting response/remission?
b.2.1.	Is a standard of care/ therapy known? (used too?) Detail
b.2.2	Is there uncertainty in the natural course of the disease (or is it possible to predict progression for each patient with 100% certainty)?
b.2.3.	The Sponsor's/Investigator's justification for a non-RCT, if any. Any other possible reasoning?
b.2.4.	If SAT(s) then how many SATs? Concomitantly conducted? Reasoning? Detail

Annex 1 (Cont.). List of characteristics used to build the studies profile

#	Characteristic
c.1.1.	Number of subgroups/strata?
c1.2.	Mutually exclusive subgroups/strata?
c.1.3.	Subgroups of equal size?
c.1.4.	Randomisation? Randomised?
c.1.5.	Two treatment arms?
c.2.1.	PIP needed/waived/deferred. Is it applicable to the setting? PIP subject to Conditional/Exceptional approval?
c.2.2.	Belief in prior? Justification Guide how much evidence is needed from what sources
c.2.3.	What is your posterior probability that there is indeed a relevant positive treatment effect, after adding evidence to your prior Guide whether or not to continue with another trial or searching for other alternative sources of evidence.
c.2.4.	Same underlying mechanism of action, similarity of response to treatment, similar dose-response relationship to conclude the mechanism is translatable to the target population?
c.2.5.	Same disease symptoms in adults and children, regarding similarity of disease progression? Determine whether full or partial extrapolation can be done.
c.2.6.	How is the timing of the paediatric trial compared to the adult trial? Detail Subsequent/in parallel/overlapping?
c.2.7.	Repurposed drug or extension of indication? If completely new drug there cannot be much confidence on extrapolation, no full extrapolation without PIP.
c.2.8.	Prior effect size estimate?
d.1.1.	Repeated measurements?
d.2.1.	More than one PEP?
d.2.2.	Co-primary endpoints? Efficacy expected in one of them?
d.2.3.	One test per endpoint?
d.2.4	Hierarchical testing?
d.2.5.	Co-primary endpoints need to be tested sequentially according to a pre-defined order/ranking? Detail
d.2.6.	Number of co-primary endpoints
d.3.1.	2, 3, 4, 5 binary endpoints? Detail
d.3.2.	Small sample sizes (1 to 50 per group)?
d.3.3.	A priori (optimal) rejection region (defined)?
d.3.4.	Prior distribution of effect sizes? Detail
d.3.5.	Power averaged over the prior distribution effect sizes?
d.3.6.	Correction method? Detail

Annex 1 (Cont.). List of characteristics used to build the studies profile

#	Characteristic
e.1.1.	Are the treatment arms in all studies the same?
e.1.2.	Is the endpoint dichotomous?
e.1.3.	Number of studies?
e.1.4.	Are all studies of equal size?
e.2.1.	Are there at least two randomized, controlled trials available?
e.2.2.	Same PEP? PEP in one of the trials used as key secondary in the other trial(s)? PEP/key secondary endpoint in one of the trials used as co-primary in the other trial(s)?
e.2.3.	Sparse events?
e.2.4.	Supportive studies similar with the pivotal trial?
e.2.5.	Treatment effect size estimate? Provided? Clearly? Detail
e.2.6.	Are the patients equally allocated per study?
f.1.1.	GAS used?
f.1.2.	Primary endpoint relevant for the entire array of patients?
f.1.3.	Heterogeneous disease course/heterogeneous population or unstable baseline?
f.1.4.	Any relevant PRO (i.e. HRQoL)? Mixed with clinician input? Carer input? Detail
f.1.5.	Details about validation (validated PRO?) Detail
f.1.5.	Patient involved in the design of PRO? Designed by clinician/patient per individual?
f.1.6.	Is the measurement at functional level relevant?

Annex 2. Data extraction form for EPARs including condition summary and criteria list

Method group X (A, B, C or D): list name [Innovative trial designs, level of evidence, study endpoints and statistical analysis, and meta-analysis]

[Name of OMP] EPAR [list number as the ASTERIX ID]

Summary of condition and summary of orphan product

[…]

Summary main clinical trials

- **(Co-)Primary endpoint(s) [list]**
- **Key secondary endpoints [list]**
- **Overall patient exposure [list number]**
- **Randomised/enrolled number patients [list number]**

Summary of how applicable the methods are and what are the adjustments, where it is the case.

OMP ID	Method	Applicable* [with/without adjustments]	Reason [list reasons] [list	Applied*	If not fully applied [list reasons]	Advantages of applying the method	Disadvantages of applying the method
[list ASTERIX OMP ID]	Method 1						
	Method 2						
	Method 3						
	Method …						

Study profile tailored around the methods and developmental plan context.
[Insert criteria list here]

Arguments
Group A, B, C or D, method 1: [list method]
[Representative extract from EPAR relevant for the applicability]
Group A, B, C or D, method 2: [list method]
[Representative extract from EPAR relevant for the applicability]

[Conclusion/Discussion on applicability of methods group X on this EPAR]

[Added value of method application]

Annex 3. Specific information for OMPs investigation extracted from the EMA documents

Subject	No	Reference number	Applicability to OMPs investigation	Reference to rare diseases	Type of studies	Study design	Endpoints	Patients Selection	Statistic considerations	Safety considerations	Paediatric considerations	Post-marketing data
General	1	EMA/CPMP/ICH/2711/99 *[ICH E11(R1)]*	General	+	+	+	+	+	+	+	+	+
	2	CPMP/ICH/375/95 *(ICH E1)*	General									
	3	CPMP/ICH/377/95 *(ICH E2A)*	General									
	4	EMA/CHMP/ICH/309348/2008 *(ICH E2F)*	General									
	5	CPMP/ICH/137/95 *(ICH E3)*	General									
	6	EMA/CHMP/ICH/435506/2012 *(Q & A for ICH E3)*	General									
	7	CPMP/ICH/378/95 *(ICH E4)*	General									
	8	CPMP/ICH/289/95 *[ICH E5 (R1)]*	General									
	9	CPMP/ICH/5746/03 *[Q & A for ICH E5 (R1)]*	General									
	10	EMEA/CHMP/EWP/692702/2008 *[ICH E5 (R1)*	General									
	11	EMA/CHMP/ICH/135/1995 *[ICH E6(R2)]*	General									
	12	CHMP/ICH/379/95 *(ICH E7)*	General	+				+				
	13	CHMP/ICH/604661/2009 *(Q & A for ICH E7)*	General									
	14	EMA/CHMP/778709/2015	General									
	15	CPMP/ICH/291/95 *(ICH E8)*	General									
	16	CHMP/EWP/2998/03/Final	General									
	17	EMA/CHMP/277591/2013	General									
	18	EMA/PDCO/CAT/CMDh/PRAC/CHMP/261500/2015	General	+	+							+
	19	EMEA/CHMP/EWP/139391/2004	General									
Biostatistics	20	CHMP/EWP/83561/2005	Specific	+	+	+	+		+		+	
	21	EMA/CHMP/295050/2013	General									
	22	CPMP/EWP/2330/99	General									
	23	EMEA/CPMP/EWP/2158/99	General	+					+			
	24	EMEA/CHMP/EWP/5872/03 Corr	General									
	25	CPMP/ICH/364/96 *(ICH E10)*	General	+		+			+			
	26	CPMP/ICH/363/96 *(ICH E9)*	General	+		+			+			
	27	EMA/CHMP/539146/2013	General									
	28	CHMP/EWP/2459/02	General	+	+	+	+		+			
	29	EMA/CPMP/EWP/1776/99 Rev. 1	General									
	30	EMA/CHMP/44762/2017	General									
	31	CPMP/EWP/482/99	General									
Clinical Pharmacology & Pharmacokinetics	32	CHMP/EWP/89249/2004	General									
	33	EMA/CHMP/83874/2014 (Rev.1)	General									
	34	CPMP/EWP/2339/02	General									
	35	EMA/618604/2008 Rev. 13	General									
	36	3CC3a	General									
	37	EMA/CHMP/458101/2016	General									
	38	CHMP/EWP/185990/06	General									
	39	EMEA/CHMP/EWP/147013/2004 Corrigendum	General									

Annex 3 (Cont.). Specific information for OMPs investigation extracted from the EMA documents

Subject	No	Reference number	Applicability to OMPs investigation	Reference to rare diseases	Type of studies	Study design	Endpoints	Patients Selection	Statistic considerations	Safety considerations	Paediatric considerations	Post-marketing data
	40	EMEA/CHMP/SWP/28367/07 Rev. 1	General									
	41	EMA/CHMP/37646/2009	General									
Blood	42	EMA/CHMP/153191/2013	Specific	+	+	+	+	+	+	+	+	+
	43	CPMP/BPWG/220/02	Specific	+	+	+	+	+		+	+	+
	44	CPMP/BPWG/2220/99	Specific	+	+	+	+	+		+	+	+
	45	CPMP/BPWP/144552/2009 Rev. 1, Corr. 1	Specific	+	+	+	+	+		+	+	+
	46	EMA/CHMP/BPWP/144533/2009 rev. 1	Specific	+	+	+	+	+		+	+	+
	47	EMA/CHMP/BPWP/153137/2011	Specific	+								
	48	CPMP/EWP/197/99	Specific	+	+	+	+	+		+		+
Cardiovascular system	49	EMEA/CHMP/EWP/356954/2008	Specific	+	+	+	+	+	+	+		+
	50	EMA/CHMP/213972/2010	Specific	+	+	+	+	+		+	+	
	51	EMA/CHMP/707532/2013	Specific	+	+	+	+	+		+	+	
	52	EMA/CHMP/494506/2012	Specific	+	+	+	+	+		+	+	+
	53	EMA/CHMP/206815/2013	Specific	+	+	+	+	+	+	+	+	+
Nervous system	54	EMA/531686/2015, Corr.1	Specific	+	+	+	+	+	+	+	+	
	55	EMA/CHMP/236981/2011, Corr. 1	Specific	+	+	+	+	+		+	+	+
	56	CHMP/EWP/566/98 Rev.2/Corr	Mixed	+	+	+	+	+	+	+	+	+
	57	EMA/CHMP/539931/2014	Mixed	+	+/-	+/-	+/-	+/+	+/-	+/-		+/-
Respiratory system	58	EMEA/CHMP/EWP/9147/2008-corr*	Specific	+	+	+	+	+	+	+	+	
Rheumatology/ musculoskeletal system	59	EMA/CHMP/239770/2014 Rev. 2	Specific	+	+	+	+	+	+	+	+	+
	60	EMA/CHMP/51230/2013	Specific	+	+	+	+	+	+	+	+	+
Antineoplastic/ immunomodulating agents	61	EMA/CHMP/205/95 Rev. 5	General	+	+	+	+	+				
	62	EMA/768937/2012	General	+								
	63	EMA/CHMP/27994/2008/Rev.1	General									
	64	EMA/CHMP/292464/2014	General	+			+					
	65	EMEA/CPMP/EWP/569/02	Specific	+	+	+	+	+			+	+
	66	EMA/CHMP/703715/2012 Rev. 2	Specific	+	+	+	+	+		+	+	
	67	CPMP/EWP/555/95 Rev. 1	Specific	+	+	+	+	+		+		
Anti-infectives for systemic use	68	EMA/CHMP/594085/2015	General									
	69	EMA/CHMP/EWP/14377/2008 Rev. 1	Specific	+	+	+	+	+	+	+	+	+
Multidisciplinary- Paediatric – Clinical efficacy and safety	70	EMA/237265/2017	Specific	+	+	+	+	+	+	+	+	
	71	EMA/199678/2016	General	+								

Annex 4. Individually conducted simulations for the 6 examples in the due template reports

Clinical Development Plan for:	Sought indication:
4.1. Defitelio® (defibrotide)	Treatment of VOD post-HSCT
4.2. Ilaris® (canakinumab)	Treatment of cryopirine-associated periodic syndromes (CAPS)
4.3. Revestive® (teduglutide)	Treament of Short Bowel Syndrome
4.4. Soliris® (eculizumab)	Treatment of long-term enzyme replacement therapy in patients with a confirmed diagnosis of aHUS disease
4.5. Fabrazyme® (agalsidase beta)	Long-term enzyme replacement therapy in patients with a confirmed diagnosis of Fabry disease
4.6. Opsumit® (macitentan)	Treatment of pulmonary arterial hypertension

Annex 4. 1. Defitelio® (defibrotide)

Clinical Development Plan

Defitelio® (defibrotide)

Clinical development in the treatment of VOD post-HSCT

Version 1; 22/08/2017

Index

2. Introduction

2.1. Background

2.1.1. Disease and currently available alternatives

Hepatic veno-occlusive disease (VOD) is a complication of vascular origin, described in patients receiving high-dose myeloablative chemotherapy as conditioning regimens for Haematopoietic Stem Cell Transplantation (HSCT).

The pathophysiology of hepatic VOD is complex; the causative event is thought to be injury to sinusoidal endothelial cells and hepatocytes in zone 3 of the hepatic acinus as a result of the high dose chemotherapy preparative regimens used for stem cell transplant procedures. Thus, hepatic VOD is considered a regimen-related toxicity. A pro-coagulant and hypofibrinolytic state is present in hepatic VOD, perhaps representing the essential underlying mechanism of this disorder.

The clinical diagnosis of hepatic VOD usually relies on the application of validated criteria described by Seattle or Baltimore criteria (Table 1). The principal difference in Baltimore criteria is the absolute requirement for bilirubin to rise to a level of >2 mg/dL (>34 µmol/L), in conjunction with ≥2 additional clinical features of hepatic VOD. Baltimore criteria are known to identify the more advanced cases of hepatic VOD.

As a consequence of liver dysfunction, patients may have hepatorenal syndrome with sodium retention and portal hypertension. This frequently progresses to acute oliguric renal failure which occurs in approximately half of the patients with severe disease, often leading to need for dialysis.

Severe hepatic VOD, is defined as VOD in the presence of multi organ failure (MOF); either pulmonary dysfunction (with an oxygen requirement when the oxygen saturation is of <90% on room air and/or ventilator dependence), and/or renal dysfunction (defined as doubling of baseline creatinine and/or dialysis dependence), and/or encephalopathy. Severe hepatic VOD is associated with a high risk of severe morbidity and mortality. Rarely do patients with severe hepatic VOD die of their liver failure; progressive MOF with consecutive lethal renal and cardiopulmonary complications are typically the main causes of death.

Well-established risk factors for hepatic VOD include age, liver inflammation, prior abdominal irradiation, hepatic fibrosis or cirrhosis, and repetitive transplants with myeloablative conditioning regimens. The conditioning regimen is also a well-established risk factor in the pathogenesis of hepatic VOD, with cyclophosphamide, total body irradiation, and particularly busulfan being the most commonly conditioning agents associated with the onset of hepatic VOD.

No therapy for the treatment or the prophylaxis of hepatic VOD had been approved previously in the US or in Europe. Investigations with several experimental approaches using anti-thrombotic and thrombolytic agents, including prostaglandin E1 and t-PA with or without concurrent heparin, have not proven successful, leaving the management of hepatic VOD

restricted to supportive care alone. The current management of VOD is limited to supportive care, such as diuretics, analgesia, haemodialysis and mechanical ventilation.

Table 1 Clinical criteria to diagnose VOD

Seattle Criteria [McDonald, 1984 Hepatology]	Baltimore Criteria [Jones, 1987 Transplantation]
Presence before day 20 after haematopoietic stem cell transplantation (HSCT) of two or more of the following: • Bilirubin ≥ 2 mg /dl (≈ 34 µmol / l) • Hepatomegaly, right upper quadrant (RUQ) pain • Ascites +/- unexplained weight gain of >2% baseline	Hyperbilirubinaemia ≥ 2 mg /dl before day 21 after HSCT and at least two of the following: • Hepatomegaly (usually painful) • Ascites • Weight gain ≥5% from baseline

2.1.2. Rationale for the development

Defibrotide is a sterile, aqueous, concentrate for solution for infusion in clear 2.5 ml glass vials. Defibrotide is a mixture of oligonucleotides obtained from porcine intestinal mucosa and prepared by controlled depolymerisation of deoxyribonucleic acid (DNA). Defibrotide had been approved for thrombotic vascular disease indications in Italy, as both ampoules for injection and oral capsules.

2.2. Scope of development

The product was developed and authorized for the following indication:

- Defitelio is indicated for the treatment of hepatic veno-occlusive disease (VOD) also known as Sinusoidal Obstructive syndrome (SOS) in haematopoietic stem-cell transplantation therapy.

2.2.1. Target product profile

Indication	Defitelio is indicated for the treatment of hepatic veno-occlusive disease (VOD) also known as Sinusoidal Obstructive syndrome (SOS) in haematopoietic stem-cell transplantation therapy.
Route of administration	IV
Pharmaceutical form	Solution for infusion containing 80 mg of defibrotide per mL in clear glass type I vials. Single use glass vial of 2.5 mL sealed with a rubber stopper
Posology	
Main target population	Patients with severe hepatic veno-occlusive disease (VOD) following haematopoietic stem-cell transplantation therapy (HSCT)
Claims to be supported by the clinical development.	Improved resolution of signs/symptoms and reduced complications (multi organ failure (MOF)) over best standard of care (BSC) in the treatment of severe hepatic veno-occlusive disease (VOD) in patients undergoing haematopoietic stem-cell transplantation therapy (HSCT)
Regions where the product should be marketed:	Global (or EU + USA)
Regulatory agencies that will be involved	EMA FDA

3. General investigational plan

3.1. Objective (s) of the development

The objective of the development is to provide pivotal support to the application for marketing authorisation in the EU of the product, by generating:

- Confirmatory evidence of (dose-response) AND/OR (superiority to best standard of care (BSC) in the treatment of VOD in HSCT.
- An appropriate safety database, including enough information to permit a characterization of the safety profile of the product, and a risk-benefit assessment of the product at the time of assessment of the marketing authorisation application in the EU.
- Any needed supporting information on the product to allow a proper risk-benefit assessment of the product at the time of assessment of the new drug marketing authorisation application in the EU.

4. Assessment of applicability of methods

4.1. Representativness of Defitelio within the cluster

Defitelio has been chosen as a potentially representative example of Single Acute episodes within the Asterix clustering of medical conditions. Some aspects, however, should be considered regarding this example that may differ from other conditions in the cluster:

- This is an acute emergency and rare complication of HSCT that implies a long-lasting enrollment and important logistic difficulties to the study conduct. Recruitment pattern is not at a constant intake rate over time as patients are not readily available for enrollment, which is a typical characteristic of these conditions and a high patient dispersion is a common denominator.
- There are not effective standard of care (SOC) treatments available. Given the seriousness of the condition, current treatment consists of supportive therapies, thus making placebo in theory acceptable on top of background SOC.
- The mechanistic rationale behind the actual development is weak (making the use of placebo reasonable option). In spite of that, there is an extensive clinical practice use throughout the patients named program which adds some practical constrains to the placebo use.
- Defibrotide had been in the Italian market for several years for the treatment of peripheral vascular conditions and there was an extensive patients named program in VOD, which a priori provides some supportive safety evidence.

4.2. Applicability of novel methodologies based on UMCU report

Applicable methodologies

- Sample size reassessment and hypothesis testing in adaptive survival trials Phase II/III (seamless)

Might be applicable

- Sequential designs for small samples
- Bayesian sample size re-estimation using power priors
- Dynamic borrowing through empirical power priors that control type I error
- Fallback-test for co-primary
- Optimal exact tests for multiple binary endpoints

Limited or non-applicability of the method

- Long-short term outcome
- Evidence, eminence and extrapolation

- Meta-analysis
- Delayed start randomization

Additional considerations

- Given the rarity of the condition, difficulties in recruitment and short period to response assessment, adaptions and sequential approaches in general may optimise the trial size. RCT with sample size reassessment and hypothesis testing in adaptive survival (phase II/III seamless) might be a suitable option. A **sequential design for small populations** may be an efficient approach. Arm-dropping adaptions may be useful if coupled to sequential approaches to support dose-finding through futility interim assessments. The applicability of both methods requires modifications to the actual conducted pivotal study.
- Bayesian methods to discard minimum required activity to continue an arm are good options for implementation of such adaptions. Bayesian approaches could be feasible also for implementing sample size reassessment or other adaptions, or for better integration of data increasing information and improving interpretation at the end of any study, especially if information on the condition is already existing and similar to the one obtained in the trial. In this particular case pre-existing data were limited and of poor quality, thus, questioning applicability in this particular case but not necessarily in other cases within the cluster.
- Methods for multiple end-points like the Fallback tests for co-primary and the optimal exact test for binary endpoints may not add value unless multiple organs are involved in heterogeneous way.
- **Enrichment methods** were not applicable given the seriousness and rarity of the conditions within the cluster. In this particular case the lack of a well defined mechanistic rationale makes even more difficult selecting a particular subset of patients more prone to response.
- Conditions with a single acute episode do not allow designs **with intrasubject controls, including cross-over and c**hallenge-dechallenge-rechallenge designs.
- External controls, if robust, could provide support to single arm trials, but only if the natural history of the condition is invariably showing poor prognosis and the size of effect in treated subjects is outstanding. Rates of response based on objective parameters or death rates in a given time frame may be compared between groups, or patients be paired with similar patients in external group, but frequency and type of assessments should be standard enough as to ensure comparability is not impaired. This would have been a suitable option for these conditions, provided that adequate historical data were available.
- Longitudinal designs with repeated measures are not suitable for acute single episode conditions.

- Delayed start is not a suitable option given the seriousness, usually life threatening, of the condition.
- **Classic designs** remain a gold standard in terms of robustness, and would usually be an acceptable option provided that there is effective treatment available or the experimental treatment is given on top of SOC, then allowing placebo control studies. This would have been one of the preferred alternative options. Nevertheless, it is acknowledged that in this particular case, might have been ethically difficult to implement, given the extensive use in clinical practice of Defitelio under the individual patient´s name program.
- Minimisation and stratification (Asterix): suitable options, already applied.
- Randomisation (Asterix): optimal if applicable.

5. Actual development plan for Defitelio in the treatment of VOD

5.1. Safety and tolerability

2 Phase 1 FIH studies in healthy volunteers,; Studies (IRI 151612 OL and DFPK88 cross-over) of the safety, tolerability, PK of defibrotide after single doses (oral and IV).

1 PK/PD ECG Study: R09-1425: double-blind, randomised, cross-over single dose study in healthy volunteers: clinical and supra therapeutic dose + placebo + moxifloxacin Pharmacokinetics

5.2. Proof of activity/dose finding

PD Study HL12326: double-blind, randomised, placebo controlled study in healthy volunteers: rising IV doses of defibrotide for 3 days in 27 patients.

Dose-finding Study 99-118: open label, randomised, uncontrolled efficacy/safety dose-finding study (also PK/PD data) of defibrotide in patients with severe VOD post HSCT, after Initial IV 10mg/kg/d followed by IV 25 mg/kg/d (75patients) or 40 mg/kg/d (74 patients) in 4 divided doses (median duration tto 14 days).

5.3. Pivotal evidence

1 pivotal non-randomised phase III study (**Study 2005-01**) to demonstrate the efficacy of Defitelio vs historical control in patients with severe VOD.

5.4. Supportive confirmatory efficacy and safety data

-**Single arm trial (2006-05)** to support treatment of severe VOD indication, OL, uncontrolled, adult and pediatric patients

-Pivotal phase III trial (**2004-000592-33**) to support prevention indication, 2 arm (defitelio vs BSC), randomised, adaptive

-**DF-Compassionate use program**, safety and efficacy, OL, uncontrolled, 711 patients with VOD

-**DF-VOD, Safety**: prospective, open, randomised and controlled (68 patients VOD post-HSCT or post-QT)

5.5. Total patient exposure in the target indication

The four principal studies in severe VOD treatment and VOD prophylaxis included a total of 647 HSCT subjects who received at least one dose of defibrotide. Nearly all subjects in the treatment and prophylaxis studies received a daily dose of 25 mg/kg/day (in 4 divided doses) with the exception of 75 subjects in 99-118 who received 40 mg/kg/day.

Additionally, a total of 1129 patients received defitelio as part of the CUP, of them efficacy and safety data are available for a total of 711 patients, mostly patients with VOD after HSCT.

5.6. Study outlines

5.6.1. Study outline dose finding study

Title of study: **Study 99-118:** Defibrotide for HSCT Patients with severe VOD: A Randomised Phase II Study to Determine the Effective Dose	
Investigators (Study center):	
Studied period: First patient included on: April 2000 Last patient completed on: May 2007	**Phase of development:** Phase II
Objectives The primary objective was to determine the complete response (CR) rate of severe VOD (total bilirubin less than 2 mg/dL and resolution of MOF) treated with defibrotide in the two dose groups. CR was defined as bilirubin <2 mg/dL after initiation of defibrotide (regardless of whether treatment was ongoing or completed) **Secondary objectives:** To catalogue the possible toxicity of defibrotide used in this setting, including Grade 3-4 toxicities and all grade toxicity; To determine the effect of defibrotide on PAI-1; To investigate whether there appears to be a dose relationship between defibrotide and/or Day +100 mortality in PAI-1 levels in these patients; To determine if one dose has a trend toward a higher therapeutic effect on Day +100 survival post-HSCT in the study population; To determine the feasibility of pharmacokinetic (PK) analysis across the two dose arms in a subset of patients and generate descriptive data of the PK of defibrotide in the study population	

Design: open label, randomised, multicentre, phase 2 CT conducted in US to determine the safety and efficacy of two doses of defibrotide [25 and 40 mg/kg/day in 4 divided doses, Arms A (n=75, ITT) and B (n=74, ITT), respectively] for the treatment of severe VOD in hematopoietic HSCT patients

Number of patients, by arm: 75 patients in defibrotide 25 mg/kg/d vs 74 patients in defibrotide 40 mg/kg/d
Intended sample size:
A total of 140 patients (70 per treatment arm) will be accrued for this study
Populations for analysis:

Main inclusion criteria:
Adult and paediatric patients were defined as those with severe VOD (defined as VOD with associated renal, pulmonary and/or CNS dysfunction) or at least a 30% risk for developing severe VOD based on the Bearman model following HSCT

Main Exclusion criteria:

Test product, dose and mode of administration, duration:
For both arms, the starting dose of defibrotide on Day 1 was 2.5 mg/kg every 6 hours (Q6H) for four doses (total dose 10 mg/kg), based on baseline weight. From Day 2, the defibrotide dose was increased to 6.25 mg/kg Q6H (total dose 25 mg/kg/day; Arm A), or 10 mg/kg Q6H (total dose 40 mg/kg/day; Arm B).
Defibrotide treatment was recommended for a minimum of 14 days or until the time of complete response, progression of hepatic VOD, or unacceptable toxicity.

Reference therapy, dose and mode of administration, duration:
NA

Criteria for evaluation (Definition, timing of assessments):
Main efficacy assessment:
Complete response (CR) rate of severe VOD (total bilirubin less than 2 mg/dL and resolution of MOF) after initiation of defibrotide (regardless of whether treatment was ongoing or completed)
Secondary variables:
Toxicity of defibrotide used in this setting, including Grade 3-4 toxicities and all grade toxicity;
Effect of defibrotide on PAI-1;
Effect of defibrotide on Day +100 mortality;
Day +100 survival post-HSCT in the study population;
Descriptive PK of defibrotide in the study population.

Statistical methods
Analysis of efficacy:
Main analysis comparing proportions by Fisher's exact test. Analysis of time to CR by log-rank methods, and description of mean time to CR using Kaplan-Meier. Subgroups were analysed for ventilator and/or dialysis dependence at study entry; age, type of CST or number of prior HSCTs.
Missing data handling: Not available
Tolerability and acceptability: Descriptive only

Main results (only actual scenario)
The results on the primary efficacy variable showed that 32 patients (43%) of Arm A and 29 patients (39%) in Arm B were considered to have had a CR. The difference between the two groups was not statistically significant with a p-value of 0.7397 (Fisher exact test).Comparison of the time to CR demonstrates no difference between the two dose groups. The median time to CR for those patients who achieved a CR equaled 40.5 days and 44.0 days, respectively for Arms A and B.
Secondary explorative analyses showed

- There were no statistically significant findings for any subgroup analysis (ventilator and/or dialysis dependence at study entry; age, type of CST or number of prior HSCTs).
  Arm A children (age ≤16) may have demonstrated a slight trend to higher CR rate compared to Arm B children (64% versus 43%; p-value 0.2362 Fisher exact test); a similar trend was observed in children when defined as age ≤18 (60% versus 42%; p-value 0.2668, Fisher exact test).
  The difference in CR in those less than 16 years in Arm A compared with Arm B in favour of the lower dose suggests that either this is a chance finding in view of the small absolute numbers or that toxicity is higher at dose B in children only (as a similar trend was not seen for those over 18 years).
  In a similar manner to that for the CR analysis, multiple subgroups analyses were performed for Day+100 survival. Key summary points from these secondary explorative analyses include:
  For children (whether defined as ≤16 or ≤18 years old) Kaplan-Meier estimates for survival at day 100 were 68.2% in Arm A versus 32.5% in Arm B for the ≤16 years subgroup and 64.0% in Arm A versus 32.5% in Arm B for the ≤18 years subgroup (p-value= 0.0259; p-value= 0.0275 respectively, log-rank test). No other significant trends noted from the subgroup analyses.

5.6.2. Pivotal Study outline

Title of study: Pivotal phase III (2005-01)
Defibrotide for the Treatment of Severe **Hepatic Veno-Occlusive Disease** in Hematopoetic Stem Cell Transplant Patients: A Historically-Controlled, Multi-Center Phase 3 Study to Determine Safety & Efficacy

Investigators (Study center): multicentre, multinational

Studied period:	**Phase of development:**
The study was performed between July 2006 and November 2008 (for collection of primary outcome data).	Phase III

Objectives

Primary:

The primary objective was to demonstrate the efficacy of defibrotide in patients with severe VOD in terms of Complete Response of severe VOD (total bilirubin less than 2 mg/dL and resolution of MOF) by Day+100 post-HSCT as the primary parameter.

Secondary:

- To compare survival at 100 and 180 days following HSCT in patients receiving defibrotide to those in a Historical Control who did not receive defibrotide;
- To assess the safety of the selected dose and schedule;
- To collect and bank samples prior to and during therapy for special studies of potential serum and endothelial markers for VOD;
- To collect historical information about the treatment centres, including severe VOD treatment across hospitals and over time and number and type of transplants per year.

Design: A historically controlled, open label, multicentre, international, Phase III clinical trial to determine the safety and efficacy of 25 mg/kg/day of defibrotide for the treatment of severe VOD in HSCT patients.

Number of patients, by arm:

Intended sample size: The original trial planned to have 80 subjects in the TG and 80 in the HC. There were several amendments in the study whereby the HC group was reduced to 32 subjects and the TG was increased to 102.

Populations for analysis: The ITT set for the Treatment Group consisted of all subjects who were consented to participate in the protocol. The ITT set for the Historical Control consisted of all subjects who were selected by the MRC as having severe VOD without any protocol exclusion criteria. The ITT analysis set was used as the primary analysis set for the primary and all secondary efficacy variables.

The Per Protocol (PP) analysis set for the Treatment Group consisted of all subjects in the intent-to- treat analysis set who received at least 21 days of defibrotide therapy. The PP analysis set for the Historical Control remained the same as for the ITT group. The PP analysis set was the secondary analysis set for all efficacy analyses. The safety analysis set consisted of all HC subjects and all TG subjects who received at least 1 dose of defibrotide.

Main inclusion criteria:

1. Adult and children
2. Clinical diagnosis of VOD by Day+21 post-HSCT (defined by jaundice (bilirubin ≥2 mg/dL) and at least 2 of the following clinical findings: ascites, weight gain ≥5% above baseline weight and/or hepatomegaly (Baltimore diagnostic criteria for VOD).
3. Subject must have severe VOD, defined as VOD with multi-organ failure (MOF); MOF is defined as the presence of one or both of the following by Day+28 post-HSCT: a) renal dysfunction (serum creatinine ≥3x value on the date of admission to the HSCT unit for conditioning or ≥3x lowest value during conditioning prior to HSCT (whichever is lowest); or Cr Cl or GFR ≤40% of admission value; or dialysis dependence) or b) pulmonary dysfunction (documentation of oxygen saturation ≤90% on room air (two consecutive measurements at least one hour apart) or requirement for oxygen supplementation/ventilator dependence).

Main Exclusion criteria:

1. Pre-existing (prior to HSCT) cirrhosis
2. An alternative diagnosis for weight gain, ascites and jaundice
3. Graft-versus-host disease (GVHD) grade B or higher involving liver or gut or grade C or higher involving skin
4. Prior solid organ transplant
5. Dependent on dialysis prior to and/or at the time of HSCT
6. Dependent on oxygen supplementation prior to HSCT
7. Significant acute bleeding or hemodynamic instability
8. Requirement for the use of any medications that increase risk of hemorrhage will be excluded from the treatment group

Test product, dose and mode of administration, duration:

All patients enrolled in the treatment group (TG) received 25 mg/kg/day of intravenous defibrotide given in 4 divided doses (approximately every 6 hours) at a maximum concentration of 4 mg/mL, each infused over 2 hours. Defibrotide was recommended to be administered for a minimum of 21 days.

Thereafter treatment was continued, as circumstances allowed, until the patient was discharged from the hospital. Defibrotide administration could be held for toxicity or delayed for necessary medical/surgical interventions. If the patient required re-hospitalization, treatment with defibrotide could be reinitiated.

Reference therapy, dose and mode of administration, duration:

Not applicable. Historical control (no treatment with defitelio). The historical control group (HC) was chosen from multiple retrospective case note reviews and the number and characteristics of the patients included in the HC were amended during the trial.

Criteria for evaluation (Definition, timing of assessments): **Efficacy:** **Main efficacy assessment (definition, timing for assessment):** The final primary endpoint was incidence of complete response of severe VOD (total bilirubin less than 2 mg/dL and resolution of MOF) by Day+100 post-HSCT, regardless of whether defibrotide treatment is ongoing or completed at the time of CR, compared with the historical control group. The initial primary endpoint chosen by the Applicant was Mortality at Day 100. This amendment was made on 3rd December 2007, before the first interim analysis formally took place at the DSMB. **Secondary variables (definition, timing for assessment):** Survival rate at 100 and 180 days post-HSCT Time to CR Concordance of CR with survival Analysis of special laboratory studies. **Safety:** Percentage of Participants With Treatment-Emergent Adverse Events [Time Frame: Through 30 days from the last dose of Defibrotide]
Statistical methods **Analysis of efficacy:** For the primary end-point and secondary end-points at different time, comparison of proportions. Time to event analysis not described in EPAR. **Missing data handling:** No details available **Tolerability and acceptability:** Description of proportions
Main results (only actual scenario) The ITT population was 102 patients in the treatment group and 32 patients in the historical control. In the Per Protocol population 61 patients were included in the treatment group and 32 patients in the historical control. Safety population 102 patients were included in the treatment group and 32 patients in the historical control. For the imputed complete response rates there was a difference when the treatment group was compared to the 32 HC group (23.5% versus 9.4%) whereas no significant difference was seen when the treatment group was compared with the original HC group of 86 (23.5% versus 19.8%). Death by D+100 63/102 (61.8%) defibrotide vs 24/32 (75%) HC, diff rate -15%, p 0.034. Death by D+180 69/102 (67.6%) defibrotide vs 24/102 (75%) HC, p 0.08. Overall TEAEs, TEAEs that led to death and TE hemorrhage event in Defitelio and HC groups were: 97% vs 64%, 64% vs 100%, and 69% vs 75%, respectively.

5.7. Uncertainties/weaknesses identified

The main uncertainties of the actual development plan are related to the lack of an adequate control arm, either concurrent (optimal) or historical (second option). This raises questions on whether treatment with Defitelio is certainly mortality reducing or if the lower rates of mortality observed compared to historical data are just due to changes/improvements in

supportive treatment, or may be due to differences in the underlying risk of the studied population that had not been adjusted completely. To note, the first study with dose-finding objective took 7 years to complete recruitment; it is likely that in that period the standard of care changed in a relevant way. The second study needed 2 years to be completed.

Thus, the basis for decision making on the risk/benefit for Defitelio was weak, because the development plan was not providing robust and unbiased evidence of efficacy and safety due to lack of reliable reference for comparison, and likely also due to change of standards of care along the period of study execution.

6. Alternative developments

6.1. Scenario 1: alternative development: RCT, placebo-controlled over SOC with sequential analysis for small populations

This option was selected based on the results of the systematic exercise of applicability of methods, but also based on the clustering of conditions, reviewing the clinical characteristics of the intended therapeutic indication. For group sequential designs, the rate of CR at 100 days was based on the rate observed in the pivotal trial, about 20% in active and 10% in historical controls.

6.1.1. Safety and tolerability

Same as main development

6.1.2. Proof of activity and dose finding

Same as main development

6.1.3. Pivotal evidence

Group sequential RCT of Defitelio D1 vs placebo on top of best supportive care to demonstrate the efficacy of Defitelio in patients with severe VOD. The study will include interim analysis with stopping rules for both futility and superiority, for one or all arms.

Title of study: Pivotal phase III (2005-01) Alternative design Option 1 Defibrotide for the Treatment of Severe **Hepatic Veno-Occlusive Disease** in Hematopoetic Stem Cell Transplant Patients: a randomized placebo-controlled group-sequential multi-Center Phase 3 study.	
Investigators (Study center): multicentre, multinational	
Studied period: The study would be done during approximately 2 years, based on the time required in the original pivotal phase III study to recruit 102 patients into a single arm design.	**Phase of development:** Phase III

Objectives

Primary:

The primary objective would be to demonstrate the efficacy of defibrotide in patients with severe VOD in terms of Complete Response of severe VOD (total bilirubin less than 2 mg/dL and resolution of MOF) by Day+100 post-HSCT as the primary parameter.

Secondary:

- To compare survival at 100 and 180 days following HSCT in patients receiving defibrotide to those in a Historical Control who did not receive defibrotide;
- To assess the safety of the selected dose and schedule;
- To collect and bank samples prior to and during therapy for special studies of potential serum and endothelial markers for VOD;

Design: A randomised placebo controlled group sequential Phase III clinical trial to determine the safety and efficacy of 25 mg/kg/day of defibrotide, given as add-on to best standard of care, for the treatment of severe VOD in hematopoietic HSCT patients. Randomisation will be stratified by ventilator and/or dialysis dependence at study entry and age group (>16 years vs =<16 years).

Number of patients, by arm: The original trial planned to have 80 subjects in the TG and 80 in the HC, thus to allow roughly detecting significant differences between treatments of 15-16% assuming 9-10% CR in placebo. Since the average sample size saving that may be obtained in sequential designs is anticipated to be 30%, the expected sample size would be 56 patients per arm (112 in total). It is anticipated that the sample size may be reached in 2 years, based on the time required to recruit 102 incident patients in a previous study.

Populations for analysis: Since it is difficult to define specific subsets for the sequential analysis, the population that would be included in the main analysis would be the randomised population. Sensibility analyses would be run thereafter for primary analysis, and as main analysis for secondary variables, on an ITT set (all subjects who were consented to participate in the protocol), a Per Protocol (PP) analysis set (all subjects in the intent-to-treat analysis set who received at least 21 days of study therapy) and a safety analysis set (all TG subjects who received at least 1 dose of study treatment).

Main inclusion criteria:

1. Adult and children
2. Clinical diagnosis of VOD by Day+21 post-HSCT (defined by jaundice (bilirubin ≥2 mg/dL) and at least 2 of the following clinical findings: ascites, weight gain ≥5% above baseline weight and/or hepatomegaly (Baltimore diagnostic criteria for VOD).
3. Subject must have severe VOD, defined as VOD with multi-organ failure (MOF); MOF is defined as the presence of one or both of the following by Day+28 post-HSCT: a) renal dysfunction (serum creatinine ≥3x value on the date of admission to the HSCT unit for conditioning or ≥3x lowest value during conditioning prior to HSCT (whichever is lowest); or Cr Cl or GFR ≤40% of admission value; or dialysis dependence) or b) pulmonary dysfunction (documentation of oxygen saturation ≤90% on room air (two consecutive measurements at least one hour apart) or requirement for oxygen supplementation/ventilator dependence).
4. All patients will receive best standard of care as background therapy, on top of which the randomized treatments will be administered.

Main Exclusion criteria:

1. Pre-existing (prior to HSCT) cirrhosis
2. An alternative diagnosis for weight gain, ascites and jaundice
3. Graft-versus-host disease (GVHD) grade B or higher involving liver or gut or grade C or higher involving skin
4. Prior solid organ transplant
5. Dependent on dialysis prior to and/or at the time of HSCT
6. Dependent on oxygen supplementation prior to HSCT
7. Significant acute bleeding or hemodynamic instability
8. Requirement for the use of any medications that increase risk of hemorrhage will be excluded from the treatment group

Test product, dose and mode of administration, duration:

All patients enrolled in the treatment group (TG) will receive 25 mg/kg/day of intravenous defibrotide given in 4 divided doses (approximately every 6 hours) at a maximum concentration of 4 mg/mL, each infused over 2 hours, and for a minimum of 21 days, as circumstances allowed, until the patient will be discharged from the hospital. Defibrotide administration could be held for toxicity or delayed for necessary medical/surgical interventions. If the patient required re-hospitalization, treatment with defibrotide could be reinitiated.

Reference therapy, dose and mode of administration, duration:

A placebo with identical appearance will be provided to be administered in a matching schedule by the IV route at the same time and volumes as determined for Defitelio, and until patient discharge as circumstances allowed.

Criteria for evaluation (Definition, timing of assessments):
Efficacy:
Main efficacy assessment (definition, timing for assessment):
Incidence of complete response of severe VOD (total bilirubin less than 2 mg/dL and resolution of MOF) by Day+100 post-HSCT, regardless of whether defibrotide treatment is ongoing or completed at the time of CR, compared with placebo
Secondary variables (definition, timing for assessment):
Survival rate at 100 and 180 days post-HSCT
Time to CR
Concordance of CR with survival
Analysis of special laboratory studies
Safety:
Percentage of Participants With Treatment-Emergent Adverse Events up to 30 days from the last dose of study treatment, TEAEs that led to death and TE hemorrhage event.

Statistical methods
Primary efficacy analysis: A group sequential analysis will be done with inspections planned after 50%, 65%, of 75% of patients will complete day 100 of follow-up. A Mantel-Haenzel test will be done for comparing the proportion of patients with CR at day 100 between groups, adjusted by stratification factors. O'Brian-Fleming adjustment for multiplicity will be applied to account for multiple inspections. Stopping rules for both futility and superiority will be determined by the result of the O'Brian-Fleming adjustment to account for a maximum level of one-sided type 1 error of 5% at the study level.
Secondary efficacy end-points at different times will be analysed similarly, and survival methods will be applied to the analysis of time to event (log-rank and Kaplan-Meier).
Missing data handling: No missing data will be imputed. A sensitivity analysis by imputing failure at the time of censoring for treatment related missingness will be done.
Tolerability and acceptability: Descriptive analysis of the proportions of TEAE, TEAEs that led to death and TE hemorrhage event.

6.1.4. Supportive confirmatory efficacy data

Same as main development

6.1.5. Supportive safety data

Same as main development

6.2. Scenario 2: alternative development: Sample size reassessment and hypothesis testing in adaptive survival studies

This option was selected based on the results of the systematic exercise of applicability of methods, but also based on the clustering of conditions, reviewing the clinical characteristics of the intended therapeutic indication. For survival designs, the time to event approach was based on the median time to CR observed in the dose-finding trial, about 40 days.

6.2.1 Safety and tolerability

Same as main development

6.2.2. Dose finding

Same as main development

6.2.3. Pivotal evidence

Parallel RCT of Defitelio D1 vs placebo on top of best supportive care to demonstrate the efficacy of Defitelio in patients with severe VOD. The study will include two interim analysis with stopping rules for both futility and superiority, and sample size reassessment.

Title of study: Pivotal phase III (2005-01) Alternative design Option 2
Defibrotide for the Treatment of Severe **Hepatic Veno-Occlusive Disease** in Hematopoetic Stem Cell Transplant Patients: a randomized placebo-controlled, adaptive survival multi-center Phase 3 Study to Determine Safety & Efficacy

Investigators (Study center): multicentre, multinational

Studied period:	**Phase of development:**
The study would be done during approximately 3 years, based on the time required in phase III study to recruit 102 patients.	Phase III

Objectives
Primary:
The primary objective will be to demonstrate the efficacy of defibrotide in patients with severe VOD in terms of time to CR as the primary parameter
Secondary:
- To compare survival at 100 and 180 days following HSCT in patients receiving defibrotide to those in a placebo;
- To assess the safety of the selected dose and schedule;
- To collect and bank samples prior to and during therapy for special studies of potential serum and endothelial markers for VOD.

Number of patients, by arm:
Intended sample size: Considering results from the previous pivotal trial, it is expected that 30 CR events will be required to detect significant differences between placebo and active, assuming a rate of CR of 9% for placebo and 23% for active in CR, and linearity of events. This corresponds to 224 enrolled patients (112 per group). Other options for sample size based on mortality require more than 500 patients per group and are deemed unfeasible.
Populations for analysis: Since it is difficult to define specific subsets during the study conduction, the population that would be included in the main analysis would be the randomised population. Sensibility analyses would be run thereafter for primary analysis, and as main analysis for secondary variables, on an ITT set (all subjects who were consented to participate in the protocol), a Per Protocol (PP) analysis set (all subjects in the intent-to-treat

analysis set who received at least 21 days of study therapy) and a safety analysis set (all TG subjects who received at least 1 dose of study treatment).

Main inclusion criteria:

1. Adult and children
2. Clinical diagnosis of VOD by Day+21 post-HSCT (defined by jaundice (bilirubin ≥2 mg/dL) and at least 2 of the following clinical findings: ascites, weight gain ≥5% above baseline weight and/or hepatomegaly (Baltimore diagnostic criteria for VOD).
3. Subject must have severe VOD, defined as VOD with multi-organ failure (MOF); MOF is defined as the presence of one or both of the following by Day+28 post-HSCT: a) renal dysfunction (serum creatinine ≥3x value on the date of admission to the HSCT unit for conditioning or ≥3x lowest value during conditioning prior to HSCT (whichever is lowest); or Cr Cl or GFR ≤40% of admission value; or dialysis dependence) or b) pulmonary dysfunction (documentation of oxygen saturation ≤90% on room air (two consecutive measurements at least one hour apart) or requirement for oxygen supplementation/ventilator dependence).

Main Exclusion criteria:

1. Pre-existing (prior to HSCT) cirrhosis
2. An alternative diagnosis for weight gain, ascites and jaundice
3. Graft-versus-host disease (GVHD) grade B or higher involving liver or gut or grade C or higher involving skin
4. Prior solid organ transplant
5. Dependent on dialysis prior to and/or at the time of HSCT
6. Dependent on oxygen supplementation prior to HSCT
7. Significant acute bleeding or hemodynamic instability
8. Requirement for the use of any medications that increase risk of hemorrhage will be excluded from the treatment group

Test product, dose and mode of administration, duration:

All patients enrolled in the treatment group (TG) will receive 25 mg/kg/day of intravenous defibrotide given in 4 divided doses (approximately every 6 hours) at a maximum concentration of 4 mg/mL, each infused over 2 hours, for a minimum of 21 days, as circumstances allowed, until the patient was discharged from the hospital.

Reference therapy, dose and mode of administration, duration:

A placebo with identical appearance will be provided to be administered in a matching schedule by the iv route at the same times and volumes as determined for Defitelio, and until patient discharge as circumstances allowed.

Criteria for evaluation (Definition, timing of assessments):

Efficacy:

Main efficacy assessment (definition, timing for assessment):

Incidence of complete response of severe VOD (total bilirubin less than 2 mg/dL and resolution of MOF) by Day+100 post-HSCT, regardless of whether defibrotide treatment is ongoing or completed at the time of CR, compared with placebo.

Secondary variables (definition, timing for assessment):

Survival rate at 100 and 180 days post-HSCT
Time to CR,
Concordance of CR with survival
Analysis of special laboratory studies.

Safety: Percentage of Participants With Treatment-Emergent Adverse Events up to 30 days from the last dose of study treatment, TEAEs that led to death and TE hemorrhage event

Statistical methods

Analysis of efficacy: An interim analysis will be done after 75% of expected events have occurred, and O'Brian-Fleming adjustment for multiplicity will be applied to account for multiple inspections. At the interim analysis, stopping rules for both futility and superiority will be determined to account for a maximum level of one-sided type 1 error of 5% at the study level. In the event of continuation, a sample size reassessment will be done.

Secondary end-points at different times will be analysed similarly, and survival methods will be applied to the analysis of time to event (log-rank and Kaplan-Meier).

Missing data handling: No missing data will be imputed. A sensitivity analysis by imputing failure at the time of censoring for treatment related missingness will be done.

Tolerability and acceptability: Descriptive analysis of the proportions of TEAE, TEAEs that led to death and TE hemorrhage event

6.2.3. Supportive confirmatory efficacy data

Same as main development

6.2.4. Supportive safety data

Same as main development

7. Analysis of the practical, ethical and regulatory impact

7.1. Analysis of option 1

Method assessed: Option 1: sequential design for small populations	Improves?	Comments
• May reduce sample size requirements	Yes	It is considered that, depending on the magnitude of the effect, group-sequential designs can achieve about 30% reduction of sample size, but may also increase study size in some circumstances
• May shorten time to study completion	Yes	Theoretically yes

• May ease recruitment	Depends	By reducing sample size may ease recruitment, but placebo control may difficult patient participation since product is already available
Statistical assessment:		
• Improves internal validity	Yes	RCT
• Increases stability of estimates	No	Probably not, in fact the decision based on the IA may be subject to less stability in the estimates
• Increases sensibility to changes	Partially	Although same variable, placebo control is more stable and theoretically enhances sensitivity to treatment-related changes
• Compliant with predetermination	Yes	By pre-established conditions for IAs
• Consistency (discuss)	No (may worsen)	It all depends on the ability to show an effect in secondary endpoints if a reduced sample size or shorter term FU. This is less a problem of this approach, if follow-up is extended after completion of main end-point
• Robustness of method (discuss)	Yes	RCT
• Protection against type I and II errors (discuss)	Yes	Multiplicity adjustments for group sequential designs for small populations, and RCT methodology
Regulatory assessment:		
• Risk of bias and credibility	Controlled	Parallel double blind design, appropriate methods for analysis
• External validity (discuss)	Similar	Similar to the previous one, trial similar size
• Therapeutic positioning and comparisons	Yes	by direct comparison vs BSC
• Informative on relevance and clinical impact	Yes	Improved by including a controlled arm
• Enough information on safety	Yes	Probably less data and shorter FU time, but this is less a problem if follow-up is extended after completion of main end-point
• Suitable information for risk-benefit balance	Yes	It can only be improved by adding a controlled arm
Ethical assessment:		
• May minimise risks	Yes	By reducing sample size
• May maximise access to treatment	No	Uncontrolled trials maximise access to treatment, but perpetuates uncertainties on whether the accessed treatment is or not efficacious

• May minimise unnecessary exposure to ineffective treatments or placebo	No	A placebo arm is introduced, but in this case there is clear equipoise
• Considers patient input	No	Hard end-points (CR based on MOF and death) are widely accepted as relevant end-points

In summary, a group sequential RCT may address uncertainties of the actual development program by providing randomised double blind evidence on causality. So, it has statistical and regulatory obvious positive consequences, at the price of including and exposing a number of patients to placebo. Although equipoise is evident even now that the product has been authorised, given the high morbid-mortality of the condition, this may raise concerns. To minimise the risk, patients will receive best standard of care, and a sequential design is proposed in order to reduce the sample size to the minimum size required to clear equipoise, and thus, to reduce exposure to placebo and/or to an ineffective treatment.

7.2. Analysis of option 2

Method assessed: Option 2: sample size reassessment and hypothesis testing in survival trials	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements	No	Survival approach requires bigger sample size
• May shorten time to study completion	No	The required sample size is bigger, so the recruitment period is the one that determines study duration
• May ease recruitment	No	Because of larger sample size
Statistical assessment:		
• Improves internal validity	Yes	Absolutely, by conducting a RCT
• Increases stability of estimates	Yes	Yes, by conducting a RCT
• Increases sensibility to changes	Yes	May adapt to deviations from actual rates in the design, allowing to correct design to account for different estimates
• Compliant with predetermination	Yes	Yes, especially considering the poor compliance in the conducted study
• Consistency (discuss)	Depends	Likely there would be better ability to show an effect in secondary endpoints
• Robustness of method (discuss)	Yes	RCT
• Protection against type I and II errors (discuss)	Yes	Adaptive survival, with protection against type I errors and room to correct design errors on sample size
Regulatory assessment:		

Method assessed: Option 2: sample size reassessment and hypothesis testing in survival trials	Improves?	Comments
• Risk of bias and credibility	Controlled	Parallel double blind design, appropriate methods for analysis
• External validity (discuss)	-	No changes
• Therapeutic positioning and comparisons	Yes	Adequate control BSC
• Informative on relevance and clinical impact	Yes	Improved by including a controlled arm
• Enough information on safety	Yes	Probably more safety data due to larger sample size
• Suitable information for risk-benefit balance	Yes	RCT, more compelling evidence will allow a better informed decision
Ethical assessment:		
• May minimise risks	No	Larger sample size, more subjects exposed to potentially ineffective treatments
• May maximise access to treatment	No	Access to the therapeutic test (i.e.: entering the trial and having chances to receive active) is reduced compared to uncontrolled studies, but larger as compared to sequential trial, since larger sample size
• May minimise unnecessary exposure to ineffective treatments or placebo	No	Increases the chances as the new design incorporates a placebo arm
• Considers patient input	No	Hard end-points (CR based on MOF and death) are widely accepted as relevant end-points

In summary, a survival adapted RCT may address the main uncertainties of the actual development program. So, it has statistical and regulatory obvious positive consequences, particularly by including randomisation and placebo control, but is less appropriate than other options from a practical point of view, since sample size requirements are larger – thus larger number of patients are required to be exposed to potentially ineffective treatments during the trial in order to clear the equipoise. Despite the equipoise is evident and the efficacy of defibrotide has not been confirmed, there may be reluctance to enter a placebo controlled trial when defibrotide is the standard treatment in clinical practice. Given the high morbid-mortality of the condition, this may raises concerns and become a barrier to study conduction. To minimise the risk, patients will receive BSC and adaptions of the design based on results of IA may allow early interruption, and thus, the exposure to placebo and/or to an ineffective treatment arm will be limited, if no longer required to conclude efficacy.

8. Recommendations

Defitelio has been chosen as a potentially representative example of Single Acute episodes within the Asterix clustering of medical conditions.

This cluster is characterised by incident cases with single acute episode, with rapid onset and rapid endpoint, well-known and predictable course in absence of treatment, and outcome often serious or life-threatening, with recovery that generally returns to baseline health status with or without sequels. Conditions in this cluster are generally led by one organ/system, although when progressing and with increasing severity multi-organic impairment may occur. The choice of control in the study design must consider whether there is an effective SOC; when a SOC is available, placebo will generally not be acceptable and then comparisons should use add-on designs. In this cluster, efficacy is based generally on a single hard objective and clinically relevant end-point, often binary.

Based on these characteristics, Defitelio in the treatment of VOD associated to HSCT is considered a representative example within the cluster of acute single episode conditions, thus the following recommendations are in general applicable to the whole cluster of conditions.

In particular, the rarity of the condition, the difficulties in recruitment and a short period to response assessment, makes in general adaptions and sequential designs suitable approaches to optimise the trial size. Placebo can be used as control treatment if added to SOC or to BSC.

In order to minimise exposure to placebo and optimise trial size, a sequential design for small populations might be the most suitable option, since has been proven more efficient and with less ethical drawbacks than an adaptive survival RCT with sample size reassessment and hypothesis testing, while both improve similarly the robustness of evidence and ability to provide causative evidence.

Bayesian methods may be applied to define stopping rules for both futility and superiority in group sequential analyses. Bayesian approaches to discard minimum required activity to continue an arm are good options for implementation of adaptions if dose finding designs are implemented such as in sequential multi-arm trials with simultaneous stopping rules, or in seamless phase II-III trials joining dose-finding with confirmation of efficacy. Bayesian approaches could be feasible also for implementing sample size reassessment or other adaptions, or for better integration of data increasing information and improving interpretation at the end of any study, especially if information on the condition already exists and is similar to the one obtained in the trial.

For disease without SOC or trials in patients who have exhausted all SOC options, controls may be historical or external; also, even uncontrolled trials assessing change from baseline or superiority to substantiated expectations may be justified, but then uncertainties are substantial and there are serious difficulties to measure size of effect and to establish causative inferences from data.

It has to be noted that in the particular case of Defitelio the extensive access throughout the patients named program made it not possible to conduct randomised placebo controlled trials. Despite this extensive use, prospective collection of data of sufficient quality for an external comparison might have solved this important limitation. However this was not done. These two conditions are directly responsible for the poor robustness of the evidence provided. Fortunately, this is an exceptional situation not necessarily linked to the cluster, and which might have been prevented with an adequate pre-planning of the clinical development program of Defitelio in the VOD indication.

Annex 4. 2. Ilaris® (canakinumab)

Clinical Development Plan

Ilaris® (canakinumab)

Clinical development in Cryopirine-associated periodic syndromes (CAPS)

Version 1; 23/09/2017

1. Index

2. Introduction

2.1. Background

2.1.1. Disease and currently available alternatives

Cryopirine-associated periodic syndromes (CAPS) are a group of monogenic auto-inflammatory diseases. CAPS is an ultra-rare (estimated incidence approximately 1:1.000.000 worldwide), autosomal dominant disease consisting of three autoinflammatory conditions of varying severity and often times overlapping symptoms [Familial cold autoinflammatory syndrome (FCAS), also called Familiar cold urticaria (FCU); Muckle-Wells syndrome (MWS); and Neonatal onset multisystem inflammatory disorder (NOMID) also known as chronic infantile neurological cutaneous and articular (CINCA) syndrome].

Although the cause is not fully clear yet, part of the etiology is the mutations in the NLRP3 gene encoding cryopirine that lead to excessive or uncontrolled IL-1β production and thus to induction of a number of inflammatory responses.

At the milder end, it is characterized by life-long, cold-induced inflammatory episodes of fever, rash and malaise (FCAS). The intermediate severity case is hereditary rheumatological syndrome of urticaria-like rash, typically associated with more intense and enduring flares and morbidity, including progressive hearing loss and kidney failure secondary to amyloidosis, associated with conjunctivitis, artralgia and fever (MWS). High mortality and nearly continuous fevers, rash, chronic aseptic meningitis, sensorineural involvement, craniofacial abnormalities and exuberant bone lesions are related with the most severe form (NOMID/CINCA).

The age of onset differs (i.e. neonatal, infancy, childhood, adolescent) and approximately 20% of untreated children with NOMID/CINCA syndrome die before reaching adulthood.

Although the majority of patients with CAPS carry mutations in the NLRP3 gene, about 50% of NOMID and 25% of MWS patients do not show such mutations but do respond to IL-1 blockade just as mutation-positive patients. Furthermore, there seems to be no clear genotype/phenotype relationship, so that the same mutation in NLRP3 may give rise to a severe phenotype in one patient and a milder phenotype in another. These findings suggest that additional so far uncharacterized genetic or environmental factors contribute to IL-1β production and initiate or modulate clinical severity.

Thus, differences in genotypes or response to treatment do not permit FCAS, MWS and NO-MID to be classed as distinct clinical entities. The earlier distinction between them is historical, and the consideration of this continuous spectrum of severities as one single disease is scientifically and medically valid. The medical literature still refers to distinct phenotypes and, consistent with a continuum of disease severity, describes overlap syndromes. Thus, a patient with symptoms of both MWS and NOMID, by the classical description, is mostly now categorized as MWS/NOMID overlap.

There are currently no approved therapies for CAPS: anakinra and various other anti-inflammatory substances are used off-label.

2.1.2. Rationale for the development

Canakinumab (ACZ885), the drug substance of Ilaris, is a human monoclonal anti-human interleukin- 1β (IL-1β) antibody of the IgG1/k isotype that was designed to bind selectively to

and neutralize the activity of IL-1β, a pro-inflammatory cytokine, which is produced mainly by mononuclear phagocytes in response to injury and infection.

2.2. Scope of development

The product is currently intended for the treatment of:

Cryopyrin-Associated Periodic Syndromes (CAPS) in adults, adolescents and children aged 4 years and older with body weight above 15 kg, including:

- Muckle-Wells Syndrome (MWS),
- Neonatal-Onset Multisystem Inflammatory Disease (NOMID) / Chronic Infantile Neurological, Cutaneous, Articular Syndrome (CINCA),
- Severe forms of Familial Cold Autoinflammatory Syndrome (FCAS) / Familial Cold Urticaria (FCU) presenting with signs and symptoms beyond cold-induced urticari-al skin rash.

2.2.1. Target product profile

Indication	Intended condition Treatment of Cryopyrin-Associated Periodic Syndromes (CAPS) in adults, adolescents and children aged 4 years and older with body weight equal to and above 15 kg, including: − Muckle-Wells Syndrome (MWS), − Neonatal-Onset Multisystem Inflammatory Disease (NOMID) / Chronic Infantile Neurological, Cutaneous, Articular Syndrome (CINCA), − Severe forms of Familial Cold Autoinflammatory Syndrome (FCAS) / Familial Cold Urticaria (FCU) presenting with signs and symptoms beyond cold-induced urticarial skin rash.
Route of administration	Subcutaneous injection
Pharmaceutical form	150 mg Powder for solution for injection
Posology	Single dose every eight weeks 150mg for patients with body weight >40kg 2mg/kg for patients with body weight <=40kg
Main target population	Adults and children aged from 4 to 17 years of age, classified as having FCAS, MWS, MWS/NOMID, or NOMID
Claims to be supported by the clinical development.	Reduction in the frequency of flares in children and adults with severe CAPS
Regions where the product should be marketed:	Global (or EU + USA)
Regulatory agencies that will be involved	EMA FDA

3. General investigational plan

3.1. Objective (s) of the development

The objective of the development is to provide pivotal support to the application for marketing authorization in the EU of the product, by generating:

- Confirmatory evidence of (dose-response) AND/OR (superiority to placebo) in CAPS.
- An appropriate safety database, including enough information to permit a characterization of the safety profile of the product, and a risk-benefit assessment of the product at the time of assessment of the marketing authorization application in the EU
- Any needed supporting information on the product to apply for a new drug authorization in EU.

4. Assessment of applicability of methods

4.1. Representativeness of Ilaris within the cluster

Ilaris has been chosen as a potentially representative example of Repeated Acute episodes within the Asterix clustering of medical conditions. Some aspects, however, should be considered regarding this example that may differ from other conditions in the cluster:

- The effect size of canakinumab is outstanding, in a condition with an extremely high rate of flaring in absence of appropriate treatment. Thus the sample size requirements to test efficacy are smaller than in other conditions, and despite the low frequency of the condition, the number of available patients for participation in clinical trials is likely to be enough to provide reasonable evidence even with conventional designs.
- No appropriate treatment is available for prevention of flares in CAPS, thus making placebo in theory acceptable
- The biological plausibility is very well established, thus acceptability of placebo in reality is small, posing ethical concerns to the delay of treatment with canakinumab.
- Canakinumab had been characterized extensively in other indications providing supportive data to the overall safety and pharmacokinetic development.

4.2. Applicability of novel methodologies based on UMCU report

Applicable methodologies:

Sample size reassessment and hypothesis testing in adaptive survival trials Phase II/III (seamless): Survival designs and adaptions are efficient in acute conditions recruiting incident cases. However, the method and the principles proposed by Magirr *et al.* regarding the drawbacks of unblinded sample size reevaluation could have been applied if the study would have been a seamless design study with a provisioned unblinded sample size re-assessment. For the applicability of this method, a continuous endpoint would have been needed, and there are doubts on the relevance of such as primary endpoint for this condition, although the methods could have been useful as there is a prior study.

Minimisation or stratification strategies: these would have been applicable based on the condition and the design of the study. This application would have plausibly offered room for improved randomisation and/or minimisation, and stratification strategies which could have been further applied to optimise the study: for instance, there was a severe imbalance between arms

(gender imbalance between placebo- and canakinumab-treated groups (93.3% females versus 6.7%)).

Sequential design for small populations: Quick response relative to recruitment. Sequential designs for small populations if the principal end-point (PEP) is replaced by a continuous endpoint might be an applicable and suitable approach, although there are doubts on the relevance of such as primary endpoint for this condition. In support of this, robust evidence for the treatment effect size was available thus this would have facilitated the power calculation and the boundaries setting.

Might be applicable:

Multi-arm group sequential designs with a simultaneous stopping rule. Taking into account that a formal dose-response study was not conducted, and given the potential applicability of sequential designs as mentioned above, then the application of a multi-arm multi-stage trial with a simultaneous stopping rule could have been possible to implement. Could have been applied to the withdrawal stage had there been more than two regimens.

Bayesian sample size re estimation using powers prior. Previous information on the clinical course can be suitable for Bayesian approaches and planning of adaptions. Bayesian approaches could be feasible in this case for implementing sample size reassessment or other adaptions, or for better integration of data increasing information and improving interpretation at the end of any study, especially if information on the condition is already existing and similar to the one obtained in the trial.

Dynamic borrowing through empirical power priors that control type I error: To be applied, good quality historical data should have been available. If two pivotal studies can be done, then dynamic borrowing could be applied to the analysis of the second trial, but gain is not anticipated to be relevant.

Not applicable:

Long-short term outcomes: there was no good biomarker available to be further validated in the phase 3 trials.

Candidate for GAS: although actually applicable, as in most chronical settings, the measurement of flares represents a reasonable and quite homogeneous clinical end-point; thus other ways to measure impact on functionality and quality of life are acceptable as a conventional support to clinical relevance.

Fallback tests for co-primary endpoints: not applicable as the clinical assessment is focused on a number of criteria for auto inflammatory flare that are all required in order to define the main end-point. Meeting one of such items in isolation is not enough as to conclude efficacy.

Optimal exact tests for multiple binary endpoints: the same as above.

4.3. Additional considerations based on disease clustering:

Regarding alternatives to the actual approach followed in the development of Ilaris, the following have been considered:

- Short period to response assessment allow adaptions and sequential approaches in general. A **sequential design based on dichotomic assessment of occurrence of first flare** may be an efficient approach to this trial. Arm-dropping adaptions may be useful if coupled to sequential approaches to support dose-finding through futility interim assessments – Bayesian methods to discard minimum required activity to continue an arm are good options for implementation of such adaptions.

- Bayesian approaches could be feasible also for implementing sample size reassessment or other adaptions, or for better integration of data increasing information and improving interpretation at the end of any study, especially if information on the condition is already existing and similar to the one obtained in the trial.
- Tools to ensure that balanced randomization is achieved are important, since the number of subjects per group is well below 30.
- Multiple end-points may not add value unless multiple organs are involved in heterogeneous way – the relatively acute nature of repeated episodes enhances the suitability of defining the "acute" event so that dichotomic and time to event approaches can be applied – this may vary within the cluster, thus more examples are desirable to test this perception.
- **Enrichment methods** are clearly improving ability to detect effect, although **may overestimate size of effect** and may be less preferable than conventional parallel designs. Parallel designs also allow to have a reference baseline to compare with, which is absent in enriched methods.
- Remitting – relapsing course is theoretically allowing **designs with intrasubject control**, the backside being that even if efficacy can be demonstrated, there **is less overall subject exposure for safety** purposes.
- Challenge-dechallenge-rechallenge methods, which were applied in the actual development for the pivotal trial, are particularly compelling and convincing to regulators, since causality conclusions derived from temporal sequence of events are robust and intuitive. Besides, from ethical perspective seem to be almost the only solution in serious conditions when a reasonable plausibility exists. However, size of effect is again overestimated, since only responders are followed – in that sense, conventional parallel designs may be regarded as less biased.
- External controls, if robust, could provide support to single arm trials, but only if the natural history of the condition is invariably showing poor prognosis and the size of effect in treated subjects is outstanding. Rates of flares may be compared between groups, or patients be paired with similar patients in external group, but frequency and type of assessments should be standard enough as to ensure comparability is not impaired.
- **Longitudinal designs with repeated measures** may be applied if number of events in a period is more important than time to first episode – in the Ilaris example, time free of flares is important, but the number of flares in a period could also be a relevant outcome. Designs where repeated measures are taken into account, with or without weighing by severity of flare, could **be an alternative for drugs with lower effect size**.
- **Delayed start can be a useful approach** in this setting, since allows to cover ethical concerns (all patients will access the experimental treatment), provides a true baseline assessment, may thus control for overestimation of effect size, and may allow to describe prognostic impact and persistence of effect on the long term.
- **Early rescue** shares benefits with delayed start, but if number of failures is anticipated to be high, may compromise blinding and trial integrity, and may difficult the estimation of the effect at the foreseen time of main assessment for efficacy, if substantially later than average rescue.
- **Classic parallel double blind randomized designs** remain a gold standard in terms of robustness, although may be logistically and ethically difficult to implement, and generally require higher sample sizes.

After these considerations, the modelling proposed has been based on a delayed start randomization for a parallel design, on one side, and a group sequential approach.

5. Actual development plan for Ilaris

The development was focused on the ability of Ilaris to reduce/prevent the number of exacerbations/flares through the long-term therapy in patients with CAPS. A summary from the EPAR is included below.

5.1. Safety and tolerability

2 Phase 1 first in man studies (A 1101 (Japanese) and B2101 (in healthy volunteers with mild asthma) of the safety, tolerability, Pk (and pharmacodynamics) of Ilaris after single (Japanese) and two doses (2-w interval).

5.2. Pharmacokinetics

Data available from 2 first in man studies (Japanese and healthy mild asthma volunteers), 2 studies in CAPS and 2 studies in different indications (psoriasis and RA). No dedicated study.

5.3. Proof of activity

A2102 (CAPS patients), to assess efficacy, safety and PK/PD of ILARIS after single dose (IV 10 mg/kg, upon relapse: 150mg sc/ or 2mg/kg 40kg).

Title of study: (Study A2102) An Open-Label, Phase II Dose Titration Study of canakinumab to Assess the Clinical Efficacy, Safety, Pharmacokinetics and Pharmacodynamics in Patients With NALP3 Mutations	
Investigators (Study center): 10 sites in 5 countries in Europe, USA and India	
Studied period: 3 ½ years Study Start Date: January 2005 Primary Completion Date: July 2008 (Final data collection date for primary outcome measure)	**Phase of development:** Phase II
Objectives **Primary:** investigate the clinical efficacy, safety, pharmacokinetics (PK) and pharmacodynamics (PD) of ACZ885, administered intravenously and subcutaneously to patients with NALP3 mutations whose clinical symptoms are either untreated or insufficiently treated and require medical intervention. **Secondary:** Dose selection	
Design: Open-label uncontrolled trial with two stages, first for dose selection starting with iv dosing, second to provide efficacy and safety data starting with iv treatment directly. For dose selection, in Stage 1, four patients received 10 mg/kg canakinumab iv infusion on Day 1, and upon relapse a second iv infusion of canakinumab (1 mg/kg), and upon second relapse, 150 mg canakinumab sc. In Stage 2, the 4 patients and additional 30 patients received 150 mg canakinumab sc injection and another 150 mg upon each relapse. The remaining patients were to provide short and long-term efficacy and safety data with the predicted dose, and received 150 mg canakinumab sc or 2 mg/kg sc if aged ≤16 years at relapse.	
Number of patients, by arm: 34 subjects, 4 in part 1 and 30 in part 2. **Intended sample size:** **Populations for analysis:**	
Main inclusion criteria: Patients aged 4 to 75 years (inclusive), Body weight ≥ 12 kg and < 100 kg.	

Females of child-bearing potential must have a negative pregnancy test. Additional birth control details to be provided at screening.
Documented molecular diagnosis of NALP3 mutations and clinical symptoms that are either untreated or insufficiently treated and require medical intervention.
Patients under anakinra therapy or any other IL-1 blocking therapy, whose clinical symptoms improved under treatment and are willing to discontinue that therapy until a relapse becomes evident.
Patients with a very severe characteristics requiring oral prednisone are eligible if the dose is stable (≤ 0.4 mg/kg/day or ≤ 20 mg/day, whichever is lower) for at least 1 week prior to the screening visit. Steroid therapy may be tapered during treatment with ACZ885 at the discretion of the investigator.
Parents' or legal guardian's written informed consent (patient's informed consent for ≥ 18 years of age) and child's assent, if appropriate, are required prior to study participation.

Main Exclusion criteria:
Participation in any clinical trial investigation (except trials with anakinra) within 4 weeks prior to dosing or longer per local regulation.
Antiinflammatory therapy with colchicine, chlorambucil, dapsone, azathioprine, mycophenolate mofetil, within 3 weeks prior to dosing. Therapeutic antibodies (e.g. anti-TNF-alpha antibodies) must be discontinued at least 60 days before dosing.
Donation or loss of 400 mL or more of blood within 8 weeks prior to dosing.
A past personal or close family medical history of clinically significant ECG abnormalities or prolonged QT-interval syndrome.
History of : Immunocompromise, including a positive HIV result ; ◦Positive Hepatitis B surface antigen or Hepatitis C test result; Drug or alcohol abuse within the 12 months prior to dosing; Tuberculosis; Renal transplant; Evidence of lymphoma.
Active medical condition preventing participation in the study such as infection, poorly controlled diabetes etc.
No live vaccinations within 3 months prior to the start of the trial, during the trial, and up to 3 months following the last dose.

Test product, dose and mode of administration, duration: Patients received canakinumab at a dose of 2 mg/kg (body weight < 40 kg) or 150 mg s.c. (body weight ≥40 kg). Patients not achieving a complete response within seven days post initial s.c. treatment received a new dose of canakinumab (5 or 10 mg/kg i.v.) as a rescue medication. Treatment including the possibility of a rescue i.v. dose was repeated upon each relapse.

Reference therapy, dose and mode of administration, duration:
Active: None
Placebo: None

Criteria for evaluation (Definition, timing of assessments): At baseline and at each study visit (post-treatment Day 1, Day 2, Week 1 and Week 5 of each treatment period), physicians assessed global disease activity and rash using a 5-point scale: absent, minimal, mild, moderate or severe. Blood samples were collected for assessment of CRP and SAA (at each study visit) and to assess hematological and biochemical markers (baseline, post-treatment Day 1, Day 2, Week 1, Week 5 and thereafter monthly each period) and immunogenicity (baseline, 1 day pre-dose, and Week 5 of each period).
Main efficacy assessment: The primary efficacy variable was time-to-relapse after achieving a complete response. A complete response was defined as a global assessment of no or minimal disease activity and no or minimal rash, and CRP and/or SAA levels within the normal range (< 10 mg/L for both parameters). Relapse was defined as having a global assessment of disease activity of mild or greater or a global assessment of disease activity of minimal and an assessment of rash of mild or greater, plus CRP and/or SAA levels of > 30 mg/L. Alternatively,

for patients with low CRP/SAA levels at baseline relapse was defined by a clinical picture necessitating retreatment, based on physician's global assessment of disease activity and rash.
Secondary variables: Secondary efficacy variables included the proportion of patients showing a complete response, physician assessments of disease activity, changes in levels of CRP and SAA.

Safety: Adverse events (AEs) were monitored throughout the study. At each study visit, physicians asked patients for, or assessed (dosing visit), any local injection site reactions following s.c. administration. Other safety assessments included the regular monitoring of vital signs, hematology blood chemistry and urinalysis, and assays for anti-canakinumab antibodiesHeight and Weight at pre-post study.
Pharmacokinetic data / Pharmacodynamic data / other: Canakinumab concentrations were assessed in serum by competitive ELISA assay (lower limit of quantification (LLOQ) = 100 ng/mL). Pharmacokinetic parameters were determined by non-compartmental analysis. IL-1β levels were analyzed.

Statistical methods
Analysis of efficacy: A Weibull gap-time frailty model was used to estimate time-to-relapse for each dose regimen. For patients who required an additional rescue dose of canakinumab, the dose regimen was defined as a separate group and the time-to-relapse was calculated from the time of i.v. rescue dose.
Missing data handling:
Tolerability and acceptability:

Main results (only actual scenario)
After the first 150 mg dose of canakinumab, 28 of 29 adult/paediatric patients achieved a complete clinical response in 2 to 9 days and 1 patient achieved a partial response. The estimated median time to relapse was 115 days, consistent with the results from the pivotal study. After the first 2 mg/kg dose, all 5 children (≤16 years, <40 kg) achieved complete clinical response within 2 to 8 days, however 2 children (both with the V198M mutation) relapsed within a week but responded to 'rescue' treatment. The median time to relapse was 48.6 days. In later periods, 3 patients again achieved complete response and 2 again needed rescue treatment.
Treatment with canakinumab resulted in a rapid onset of efficacy, with disappearance or significant improvement of symptoms within one day after dosing (day 2). Laboratory parameters such as high CRP or SAA normalized within days of canakinumab injection. Improvements in audiologic tests were noted in 6 patients and quality of life assessments showed general trends for improvement after dosing.

5.4. Dose finding

No formal dose finding was done: There is no classical therapeutic exploratory Phase II study in CAPS patients. Data from the first part (single dose) of study A2102 was combined with modelling and simulation to predict the dose and the dosing regimen used in the subsequent CAPS studies

5.5. Pivotal evidence

1 pivotal randomized withdrawal phase III study (**D2304**) to demonstrate the efficacy of ILARIS vs placebo in patients with MWS.

Title of study: Randomized withdrawal phase III study (**D2304**) to demonstrate the efficacy of ILARIS vs placebo in patients with MWS.

Investigators (Study center): multicentre, multinational – 35 patients from 11 centers in five countries (France, Germany, India, the United Kingdom, and the United States)

Studied period: 18 months Study Start Date: April 2007 Primary Completion Date: October 2008 (Final data collection date for primary outcome measure)	**Phase of development:** Phase III

Objectives
Primary: The primary objective was to assess the efficacy of canakinumab (percentage of patients who experienced disease flare) compared with placebo in Part II as determined by the Physician's global assessment of autoinflammatory disease activity, assessment of skin disease and inflammation markers (CRP and/or SAA).
Secondary: objectives were to assess the safety, tolerability and immunogenicity of canakinumab, to assess overall efficacy (response rate) of canakinumab in Part I and Part III as determined by the Physician's global assessment of autoinflammatory disease activity, assessment of skin disease and inflammation markers, to evaluate the pharmacokinetics (PK) and pharmacodynamics (PD) of canakinumab and to assess the effect on disease progression with regards to deafness, kidney function, neurological and ophthalmological symptoms.

Design: 2 arms, 3 stage, double-blind, parallel withdrawal (in stage 2), placebo-control, ND-masking, randomization, no-stratification, biomarkers: CRP & SAA (serum amyloid A).

Number of patients, by arm: A total of 41 patients were screened and 35 enrolled in Part I, of whom 31 were randomized into Part II (15 canakinumab: 16 placebo). All these 31 patients entered Part III. 35 exposed to experimental, 16 to placebo.
Intended sample size: The total sample size was planned to be at least be 20 patients randomized into the withdrawal period (Part II). The sample size was calculated to show superiority of canakinumab relative to placebo for the proportions of patients with disease flare. A sample size of 10 patients per group would give 90% power to detect a treatment difference between disease flare rate of 15% for the active group and of 90% for the placebo group, based on Fisher's exact test with a 0.05 two-sided significance level.
Populations for analysis: All patients were included in the safety and intent to treat populations.

Main inclusion criteria: Adults and children ages of 4 and 75 years and weighed at least 15 kg (33 lb) but less than 100 kg (220 lb), with CAPS associated with an NLRP3 mutation and who required treatment.

Main Exclusion criteria: use of corticosteroids ≥20 mg/day or >0.4 mg/kg for 1 week; colchicine, dapsone or mycophenolate mofetil for 3 weeks; etanercept, leflunomide, thalidomide or ciclosporin for 4 weeks; adalimumab or intravenous immunoglobulin for 8 weeks; or infliximab, 6-mercaptopurine, azathioprine, cyclophosphamide or chlorambucil for 12 weeks. Females entering the study were prepubescent, postmenopausal, sterilized or required to use effective contraception. Patients with HIV/AIDS, significant medical history of other conditions, history of recurrent infections or live vaccination within 3 months of the study, or who had an abnormality on their electrocardiogram were excluded from the study.

Test product, dose and mode of administration, duration: All patients received 1 injection of canakinumab in Part I (8 weeks) and those who had a complete and sustained response were randomized in Part II to receive a maximum of 3 injections of canakinumab or placebo at 8 week intervals. Patients who finished Part II, or flared beforehand, entered Part III and received canakinumab every 8 weeks until 48 weeks. Treatment duration 48 weeks – 8 weeks

run-in, 24 weeks randomized, 16 follow-up. Doses: 150 mg s.c. if >40kg and 2 mg/kg s.c. when ≤ 40kg, both every 8 weeks.

Reference therapy, dose and mode of administration, duration:
Active: None
Placebo: Yes, double blinding but characteristics of placebo not described – assumed to be 3 subcutaneous administrations separated by 8 weeks unless withdrawal.

Criteria for evaluation (Definition, timing of assessments):
Efficacy assessments: Physicians assessed global disease activity and each of: urticarial rash, arthralgia, myalgia, headache or migraine, conjunctivitis, fatigue or malaise, and other symptoms related or unrelated to CAPS, by a 5-point scale for disease activity: absent, minimal, mild, moderate, or severe. Patients performed a global assessment of their symptoms together with assessments of each of the following symptoms: fever or chills, rash, joint or muscle pain, eye discomfort or redness, fatigue, headache, and other symptoms, using the same 5-point scale used by physicians.
Timing of assessments: At weeks 0, 1, 2 (for patients with a partial response at week 1), 4 and 8, and 4-weekly thereafter up to week 48. IGA (physician global assessment) and blood sampling not done at week 4.
Main efficacy assessment: Proportion of patients with disease flare in Part II (defined as those who experienced a protocol-defined clinical relapse, or discontinued from Part II for any reason) compared to placebo. Disease relapse was defined as a CRP and/or SAA values greater than 30 mg/L, as well as a physician global assessment of auto-inflammatory disease that was either rated above minimal or, if equal to minimal, accompanied by skin disease activity rated above minimal (assessed on the same day).
Secondary variables: Proportion of canakinumab treatment responders in Part I; - Proportion of patients without disease relapse in Part III; Change in inflammatory markers CRP and SAA; Investigator's clinical assessment of auto-inflammatory disease activity; Patient's assessment of symptoms; Assessment of skin disease; Medical Outcome Short Form (36) Health Survey (SF-36); Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-F); Health Assessment Questionnaire (HAQ); Child Health Questionnaire – Parent Form (CHQ-PF87) (in children/adolescent patients)

Safety: Adverse events were recorded throughout the study; the severity of such events and their relationship to the administration of a study drug were recorded. Patients were asked about the occurrence of any injection-site reactions.
Pharmacokinetic data / Pharmacodynamic data / other: Blood samples were collected to measure levels of acute-phase reactants, C-reactive protein (CRP), and serum amyloid A protein (SAA) and to assess hematologic and biochemical markers and immunogenicity. Blood samples were analyzed in a central laboratory by pathologists who were unaware of study-group assignments.

Statistical methods
Analysis of efficacy: Primary analysis was based on ITT in part 2 (all randomized patients receiving at least one dose of study drug). The study groups were compared with the use of the Fisher's exact test. The time until disease flare in part 2 was assessed with the use of a Cox proportional-hazards regression model. Kaplan–Meier estimates were plotted against time. Changes in inflammatory markers from week 8 were analysed with the use of a Wilcoxon rank-sum test. All statistical tests were two-sided at a significance level of 0.05. In Parts I, II and III the proportion of patients who responded without disease relapse was calculated. In Part II treatment comparison was performed on common odds ratio by a stratified test procedure.
Missing data handling: Patients who met the criteria for relapse or who discontinued treatment prematurely in part 2 for any reason were considered to have had a disease relapse

for ITT. One patient in the canakinumab treatment group was excluded from the per protocol population due to missing primary endpoint assessments.
Tolerability and acceptability: Fisher's exact test was used to compare the incidence of infections between study groups in part 2.

Main results (only actual scenario)
Part I: 31/35 completers, complete responders 97.1%, 71.4% response from first assessment at day 8. Part II: Proportion of patients who experienced a disease flare: Canakinumab 0/15 (0%). Placebo 13/16 (81,3%). Odds ratio=0,0 (95%CI 0,0 - 0,14). p<0,001. Part III: 96.8% without relapses at the end of 16 weeks.

5.6. Supportive confirmatory efficacy and safety data

1 Supportive study roll-over open label study (**D2306**) including patients completing previous studies and also new patients.

Title of study: An Open-label, Long-term Safety and Efficacy Study of ACZ885 (Anti-interleukin-1β Monoclonal Antibody) Administered for at Least 6 Months in Patients With the Following Cryopyrin-associated Periodic Syndromes: Familial Cold Autoinflammatory Syndrome, Muckle-Wells Syndrome, or Neonatal Onset Multisystem Inflammatory Disease

Investigators (Study center): multicentre, multinational –33 sites in 9 countries in EU, USA and India

Studied period:	**Phase of development:**
Study Start Date: May 2008 Study Completion Date: April 2010 (cut-off for EPAR on Jan 9th 2009, with 98 subjects	Phase III

Objectives
Primary: To provide long-term safety and efficacy data for ACZ885 (a fully human anti-interleukin-1β [anti-IL-1β] monoclonal antibody) given as an injection subcutaneously in patients who participated in the CACZ885A2102 (NCT00487708), CACZ885D2201 (NCT00685373) or CACZ885D2304(NCT00465985) studies or newly identified patients with the following cryopyrin-associated periodic syndromes: Familial Cold Autoinflammatory Syndrome, Muckle-Wells Syndrome or Neonatal Onset Multisystem Inflammatory Disease.
Secondary:

Design: Open label uncontrolled trial

Number of patients, by arm: 166 patients, single arm – at cut-off, 98 subjects (54 from studies (24 from PoC, 30 from randomized) and 44 naïve)
Intended sample size: As determined by completion of other studies, and additional patients.
Populations for analysis: Analyses included all patients who received at least one dose of study treatment.

Main inclusion criteria: This study included participants previously enrolled in CACZ885A2102 (NCT00487708), CACZ885D2304 (NCT00465985) and ACZ885 naive patients.
1. Male and female patients at least 3 years of age
2. Diagnosis of Familial Cold Autoinflammatory Syndrome, Muckle-Wells Syndrome or Neonatal Onset Multisystem Inflammatory Disease. Prior agreement between the Investigator and Novartis for study eligibility is required for patients who do not have a molecular diagnosis of NALP3 mutations available (either testing not performed, or testing

performed but negative)upon study entry. For those patients who have not been molecularly tested for NALP3 mutations, molecular testing should be performed during the course of the study
3. For patients under anakinra therapy or any other investigational IL-1 blocking therapy, these treatments should be discontinued prior to the baseline visit.

Main Exclusion criteria:
1. Participation in any clinical investigation within 4 weeks prior to dosing or longer if required by local regulation with the exception of trials with anakinra, other investigational IL-1 blocking therapies, and/or ACZ885.
2. History of being immunocompromised, including a positive HIV at screening (ELISA and Western blot) test result.
3. A positive Hepatitis B surface antigen (HBsAg) or Hepatitis C antibody test result is not allowed.
4. No live vaccinations within 3 months prior to the start of the trial, during the trial, and up to 3 months following the last dose.
5. History of recurrent and/or evidence of active bacterial, fungal, or viral infections.
6. Positive tuberculin skin test reaction (PPD 5 tuberculin units or as according to local standard practice) (>= 5 mm induration) at 48 to 72 hours after administration at the screening visit or within 2 months prior to the screening visit. Patients who have a positive PPD skin test with a documentation of BCG vaccination, who are at low environmental risk for tuberculosis (TB) infection or reactivation, and have a negative chest X-ray can be included.

Test product, dose and mode of administration, duration: Subcutaneous injection every 8 weeks based on participant's body weight. Body weight >40 kilogram (kg): 150 milligrams (mg) per injection and body weight <= 40 kg: 2 mg/kg per injection. For participants who did not experience sufficient symptomatic relief, an up-titration to the dose and/or more frequent doses were permitted as per protocol.

Reference therapy, dose and mode of administration, duration:
Active: None. **Placebo:** None

Criteria for evaluation (Definition, timing of assessments):
Timing of assessments: Time Frame: 2 years depending on when the participant enters the study (main safety); Every 8 weeks during the course of the trial for at least 6 months with a maximum duration of 2 years (efficacy secondary).
Main safety assessment: The Number of Participants With Adverse Events (AEs), Death, Serious Adverse Events (SAEs), Discontinuation of Study Drug Due to an AE, Infections and Infestations and Injection Site Reactions.
The number of participants with Adverse Events and Infections & Infestations are regardless of study drug relationship by primary system organ class preferred term equal and/or greater than 2% in any group. The number of participants with mild injection site reactions= mild reactions observed on at least one occasion but no moderate or severe reactions. The number of participants with moderate injection site reactions= moderate reactions observed on at least one occasion but no severe reactions.
Secondary efficacy variables: The Percentage of Participants Without Disease Relapse as Determined by the Physician's Global Assessment of Autoinflammatory Disease Activity, Assessment of Skin Disease and Inflammation Markers. Disease relapse following complete response is defined as inflammation markers: C-Reactive Protein (CRP) and/or Serum Amyloid A (SAA) result > 30 mg/L AND Physician's Global Assessment of Autoinflammatory Disease Activity > minimal or Physician's Global Assessment >= minimal AND Skin Disease Assessment > minimal. Physician's Global Assessment of Autoinflammatory Disease Activity

and Skin Disease Assessment (urticarial skin rash) are completed by the investigator using a 5 point rating scale: absent, minimal, mild, moderate and severe. **Pharmacokinetic data / Pharmacodynamic data / other:** Pharmacokinetics: Mean Clearance from serum in Liter per Day (CLD) in adult participants >=18, pediatric participants <18 with body weight >40 kg and pediatric participants <18 with body weight <=40 kg. Immunogenicity of Canakinumab (ACZ885). The number of participants who tested positive for anti-ACZ885 antibodies using the Biacore Assay at the end of the study. Audiograms (timing?). Ophtalmologic assessments (timing?)
Statistical methods Descriptive statistics were used to summarise demographics, baseline characteristics, efficacy and safety. **Missing data handling:** Missing values were not imputed.
Main results (only actual scenario): **EPAR:** Up to the data cut-off (09-Jan-2009), 98 patients exposed (54 roll-over patients: 24 patients from A2102 and 30 from D2304) and 44 naïve. 150 mg s.c. was effective in 77/ 98 CAPS patients (77/88 patients with available response data) and 16 (16.3%) patients needed intensified dosing regimen. Canakinumab-naïve patients: 41/44 patients for whom response data was available at database lock achieved a complete clinical response. **Publication:** Complete response was achieved in 85 of 109 canakinumab-naive patients (78%; 79/85 patients within 8 days, and five patients between days 10 and 21). Of 141 patients with an available relapse assessment, 90% did not relapse, their CRP/SAA levels normalised (<10 mg/l) by day 8, and remained in the normal range thereafter. Median treatment duration was 414 days (29-687 days). Upward adjustments of dose or frequency were needed in 24.1% patients; mostly children and those with severe CAPS. Predominant adverse events (AE) were infections (65.7%) of mostly mild-to-moderate severity. Serious AE reported in 18 patients (10.8%) were mainly infections and were responsive to standard treatment. The majority of patients (92%) reported having no injection-site reactions and only 8% patients reported mild-to-moderate reactions. Patients receiving vaccination (15%) showed normal immune response.

5.7. Total patient exposure in the target indication

Efficacy, safety and tolerability data on canakinumab from CAPS studies (completed or with an interim data cut-off) were available for 104 subjects (including 23 children aged from 4 to 17 years of age). Of these, 20 patients were classified as having FCAS, 72 as MWS, 10 with MWS/NOMID, and 1 with NOMID. Patients were observed for up to 3¾ years, the overall duration of exposure was approximately 96 patient years.

5.8. Uncertainties/weaknesses identified

No formal dose-finding studies have been conducted, and thus it is uncertain if a different dosing regimen might have been a better option from an efficacy and/or safety point of view.

Demonstration of efficacy is limited to short-term control of inflammation and related symptoms/signs, including reductions in inflammatory markers like SAA and CRP, but the long-term benefits, i.e. reduction/prevention of amyloidosis and end organ damage caused by the inflammatory process, have not been formally demonstrated.

The pivotal trial was conducted in MWS patients, other phenotypes were included in open label studies. Induction and maintenance of response showed similarity between the different clinical phenotypes of CAPS ranging from FCAS over MWS to NOMID including overlap syndromes.

Considering the known pathogenesis and the fact that patients had confirmed mutation in the NALP3 gene, inclusion of these patients in the indication appeared justified.

No patients without confirmed genetic mutation were included in the trials. Preliminary data indicated that phenotypically indistinguishable patients may benefit from treatment as well.

The safety database was limited at the time of authorization: 56 patients had been treated for more than 48 weeks, only 6 patients had been treated for more than 96 weeks. Thus long term safety beyond 1 year was not well known.

Efficacy has also been demonstrated in children; however duration of treatment response appears to be shorter in some patients, a finding that should be investigated further in combination with further PK/PD studies.

6. Alternative development plans

6.1. Option 1: alternative development: Parallel delayed start pivotal study

This option was selected based on the results of the systematic exercise of applicability of methods.

6.1.1. Safety and tolerability

Same as main development

6.1.2. Proof of activity/dose finding

Same as main development.

6.1.3. Confirmatory efficacy data

A randomized double-blind delayed start comparison of canakinumab and placebo in the prevention of flares in patients with CAPS

Title of study: A randomized double-blind delayed start comparison of canakinumab and placebo in the prevention of flares in patients with CAPS	
Investigators (Study center): multicenter, minimum 5 countries, 11 sites.	
Studied period: 6 months recruitment, follow-up 16 weeks (main phase) + 32 weeks (maintenance) First patient included on Last patient completed on	**Phase of development:** Phase III
Objectives **Primary:** to assess the efficacy of canakinumab (percentage of patients who experienced disease flare) compared with placebo as determined by the Physician's global assessment of autoinflammatory disease activity, assessment of skin disease and inflammation markers (CRP and/or SAA). **Secondary:** to assess the safety, tolerability and immunogenicity of canakinumab, to assess the effect on the course of the condition of a delayed start of the treatment on audiometry, ophtamological and neurological assesments, and renal impairment (?).	

Design: Stage 1 randomised double blind treatment with placebo and canakinumab every 8 weeks (based on phase I PK data). If increase of inflammatory signs or symptoms before 16 weeks the patient will be rescued with one dose of active treatment and is switched to second stage. If no response is achieved with rescue, the patients is dropped from the study as a failure and treated according to investigator criteria. Second stage will be open-label treatment with canakinumab up to 32 additional weeks (up to 48 weeks).

Number of patients, by arm: 30 patients, 15 per arm
Intended sample size:
The original sample size assumed rates of 15% for placebo and 90% for active in the randomised withdrawal phase. Assuming a more modest effect, 30 patients (15 per arm) will be needed to detect a difference of 50% between treatments or higher with ≈90% power and 5% two-sided protection against type I error, assuming a 85% relapse rate in placebo and 25% in canakinumab. Sample lack of eligibility would account for 20%, requiring to screen up to 38 patients.
Populations for analysis: ITT (as assigned)

Main inclusion criteria: Adults and children ages of 4 and 75 years and weighed at least 15 kg (33 lb) but less than 100 kg (220 lb), with CAPS associated with an NLRP3 mutation and who required treatment, and who have been receiving estable anti-inflammatory or immunosuppressant treatment in the last weeks, as detailed in exclusion criteria, and with estable clinical symptoms in the last week.

Main Exclusion criteria: use of corticosteroids ≥20 mg/day or >0.4 mg/kg for 1 week; colchicine, dapsone or mycophenolate mofetil for 3 weeks; etanercept, leflunomide, thalidomide or ciclosporin for 4 weeks; adalimumab or intravenous immunoglobulin for 8 weeks; or infliximab, 6-mercaptopurine, azathioprine, cyclophosphamide or chlorambucil for 12 weeks. Females entering the study were prepubescent, postmenopausal, sterilized or required to use effective contraception. Patients with HIV/AIDS, significant medical history of other conditions, history of recurrent infections or live vaccination within 3 months of the study, or who had an abnormality on their electrocardiogram were excluded from the study.

Test product, dose and mode of administration, duration:
2 injections of canakinumab or placebo at 8 week intervals. Patients who finished stage 1 or are rescued will receive canakinumab every 8 weeks until 48 weeks.

Reference therapy, dose and mode of administration, duration:
Active:
Placebo: Yes, undistinguishable from canakinumab

Criteria for evaluation (Definition, timing of assessments):
Efficacy assessments: Physicians assessed global disease activity and each of: urticarial rash, arthralgia, myalgia, headache or migraine, conjunctivitis, fatigue or malaise, and other symptoms related or unrelated to CAPS, by a 5-point scale for disease activity: absent, minimal, mild, moderate, or severe. Patients performed a global assessment of their symptoms together with assessments of each of the following symptoms: fever or chills, rash, joint or muscle pain, eye discomfort or redness, fatigue, headache, and other symptoms, using the same 5-point scale used by physicians.
Timing of assessments: At weeks 0, 1, 2 (for patients with a partial response at week 1), 4 and 8, and 4-weekly thereafter up to week 48. IGA (physician global assessment) and blood sampling not done at week 4.
Main efficacy assessment: Proportion of patients with disease flare in Part II (defined as those who experienced a protocol-defined clinical relapse, or discontinued from Part II for any reason) compared to placebo. Disease relapse was defined as a CRP and/or SAA values greater than 30 mg/L, as well as a physician global assessment of auto-inflammatory disease that was

either rated above minimal or, if equal to minimal, accompanied by skin disease activity rated above minimal (assessed on the same day).
Secondary variables: Proportion of patients without disease relapse in stage 2; Change in inflammatory markers CRP and SAA; Investigator's clinical assessment of auto-inflammatory disease activity; Patient's assessment of symptoms; Assessment of skin disease; Medical Outcome Short Form (36) Health Survey (SF-36); Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-F); Health Assessment Questionnaire (HAQ); Child Health Questionnaire – Parent Form (CHQ-PF87) (in children/adolescent patients)

Safety: Adverse events were recorded throughout the study; the severity of such events and their relationship to the administration of a study drug were recorded. Patients were asked about the occurrence of any injection-site reactions.
Pharmacokinetic data / Pharmacodynamic data / other: Blood samples were collected to measure levels of acute-phase reactants, C-reactive protein (CRP), and serum amyloid A protein (SAA) and to assess hematologic and biochemical markers and immunogenicity. Blood samples were analyzed in a central laboratory by pathologists who were unaware of study-group assignments.

Statistical methods
Analysis of efficacy: Primary analysis based on ITT in stage1 (all randomized patients receiving at least one dose of study drug). Study groups comparisons by the Fisher's exact test. The time until disease flare in stage 1: Cox proportional-hazards regression model. Kaplan–Meier estimates plotted against time. Changes in inflammatory markers from week 8: Wilcoxon rank-sum test. All statistical tests two-sided at a significance level of 0.05.
In stage 1 and 2 the proportion of patients who responded without disease relapse will be calculated.

Missing data handling: Patients who met the criteria for relapse or who discontinued treatment prematurely in any stage for any reason will be considered to have had a disease relapse for ITT.
Tolerability and acceptability: Fisher's exact test to compare the incidence of infections between study groups in part 1.

6.1.4. Supportive safety data

Same as main development

6.2. Option 2: Multi-arm group sequential design

This option was selected partly based on the results of the systematic exercise of applicability of methods, suggesting that sequential methods are applicable. However, both multi arm multistage sequential trials with simultaneous stopping rules and sequential designs for small samples require continuous variables to be applied. Thus, an adaptive design comparing two doses versus placebo in a sequential manner with the possibility to drop an ineffective dose for binary variables is proposed. Interim analyses are planned at 60% and 80% of patients for futility (for the worst dose and for the study), and overwhelming superiority of any dose. Conclusion of utility may lead to interruption of one of the active dose arms at any interim.

6.2.1. Safety and tolerability

Same as main development

6.2.2. Proof of activity

Same as main development

6.2.3. Dose finding/pivotal study

Sequential trial with multiple arms testing more than one dose level, based on parallel design with adaptive methods aimed to drop the worse dose based on futility.

Title of study: A randomized double-blind group sequential dose-finding and efficacy confirmatory trial testing two doses of canakinumab and placebo in the prevention of flares in patients with CAPS	
Investigators (Study center): multicenter, minimum 5 countries, 11 sites.	
Studied period: 6 -12 months recruitment, follow-up 16 weeks (main phase) and then up to 48 weeks. Total 18-24 months	**Phase of development:** Phase III
Objectives **Primary:** to assess the efficacy of canakinumab (percentage of patients who experienced disease flare) compared with placebo as determined by the Physician's global assessment of autoinflammatory disease activity, assessment of skin disease and inflammation markers (CRP and/or SAA). **Secondary:** to assess the safety, tolerability and immunogenicity of canakinumab, to assess the effect on the course of the condition of a delayed start of the treatment on audiometry, ophthalmological and neurological assessments, and renal impairment (?).	
Design: Stage 1 randomised double blind treatment with placebo or canakinumab every 8 weeks (based on phase I PK data). If increase of inflammatory signs or symptoms before 16 weeks the patient will be rescued with one dose of active treatment and is switched to second stage. If no response is achieved with rescue, the patients is dropped from the study as a failure and treated according to investigator criteria. Second stage will be open-label treatment with canakinumab up to 32 additional weeks (up to 48 weeks). No stratification is planned.	
Number of patients, by arm: up to 45 patients, 15 per arm **Intended sample size:** up to 45 patients (15 per arm) to detect a difference of 50% between treatments or higher with 90% power and 5% protection against type I error, assuming a 85% relapse rate in placebo and 35% in canakinumab. Sample lack of eligibility would account for 20%, requiring to screen up to 57 patients (19 per group). Since in average sequential trials allow for 30% sample size saving, there would be chances to finish with similar sample size as initially planned. Interim analyses are planned at 27 patients (60%) and 36 patients (80%), while reducing risk for a negative trial. Finally, a sample size reassessment is planned at the second interim analysis to cope with potentially deviation in estimates from variability in the study design. **Populations for analysis:** ITT (as assigned)	
Main inclusion criteria: Adults and children ages of 4 and 75 years and weighed at least 15 kg (33 lb) but less than 100 kg (220 lb), with CAPS associated with an NLRP3 mutation and who required treatment, and who have been receiving stable anti-inflammatory or immunosuppressant treatment in the last weeks, as detailed in exclusion criteria, and with stable clinical symptoms in the last week.	
Main Exclusion criteria: use of corticosteroids ≥20 mg/day or >0.4 mg/kg for 1 week; colchicine, dapsone or mycophenolate mofetil for 3 weeks; etanercept, leflunomide,	

thalidomide or ciclosporin for 4 weeks; adalimumab or intravenous immunoglobulin for 8 weeks; or infliximab, 6-mercaptopurine, azathioprine, cyclophosphamide or chlorambucil for 12 weeks. Females entering the study were prepubescent, postmenopausal, sterilized or required to use effective contraception. Patients with HIV/AIDS, significant medical history of other conditions, history of recurrent infections or live vaccination within 3 months of the study, or who had an abnormality on their electrocardiogram were excluded from the study.
Test product, dose and mode of administration, duration: 2 dose regimens for canakinumab, with required dummy to be undistinguishable externally and when administered. 2 injections of canakinumab or placebo at 8 week intervals. Patients who finished stage 1 or are rescued will receive canakinumab every 8 weeks until 48 weeks.
Reference therapy, dose and mode of administration, duration: **Active:** None **Placebo:** Yes, undistinguishable from canakinumab externally and when administered.
Criteria for evaluation (Definition, timing of assessments): **Efficacy assessments:** Physicians assessed global disease activity and each of: urticarial rash, arthralgia, myalgia, headache or migraine, conjunctivitis, fatigue or malaise, and other symptoms related or unrelated to CAPS, by a 5-point scale for disease activity: absent, minimal, mild, moderate, or severe. Patients performed a global assessment of their symptoms together with assessments of each of the following symptoms: fever or chills, rash, joint or muscle pain, eye discomfort or redness, fatigue, headache, and other symptoms, using the same 5-point scale used by physicians. **Timing of assessments:** At weeks 0, 1, 2 (for patients with a partial response at week 1), 4 and 8, and 4-weekly thereafter up to week 48. IGA (physician global assessment) and blood sampling not done at week 4. **Main efficacy assessment:** Proportion of patients with disease flare in Part 1 (defined as those who experience a protocol-defined clinical relapse, or discontinue from Part 1for any reason) compared to placebo. Disease relapse is defined as a CRP and/or SAA values greater than 30 mg/L, as well as a physician global assessment of auto-inflammatory disease, either rated above minimal or, if equal to minimal, accompanied by skin disease activity rated above minimal (assessed on the same day). **Secondary variables:** Proportion of patients without disease relapse in stage 1; time to flare; Change in inflammatory markers CRP and SAA; Investigator's clinical assessment of auto-inflammatory disease activity; Patient's assessment of symptoms; Assessment of skin disease; Medical Outcome Short Form (36) Health Survey (SF-36); Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-F); Health Assessment Questionnaire (HAQ); Child Health Questionnaire – Parent Form (CHQ-PF87) (in children/adolescent patients),Proportion of patients without disease relapse in stage 1 **Safety:** Adverse events recorded throughout the study; severity and their relationship to the administration of a study drug will be recorded. Patients will be asked about the occurrence of any injection-site reactions. **Pharmacokinetic data / Pharmacodynamic data / other:** Blood samples will be collected to measure levels of acute-phase reactants, C-reactive protein (CRP), and serum amyloid A protein (SAA) and to assess hematologic and biochemical markers and immunogenicity. Blood samples will be analysed in a central laboratory by pathologists unaware of study-group assignments.
Statistical methods **Analysis of efficacy:** Primary analysis at week 16 based on ITT (all randomized patients receiving at least one dose of study drug) and study groups comparisons by the Fisher's exact test. Two Interim analyses are planned at 27 patients (60%) and 36 patients (80%). O'Brian-

Fleming adjustment for multiplicity will be applied to account for multiple inspections. Stopping rules for both futility and superiority will be determined by the result of the O'Brian-Fleming adjustment to account for a maximum level of one-sided type I error of 5% at the study level. At any inspection, arms may be interrupted for futility or superiority to placebo, based on predetermined criteria. Since in average sequential trials allow for 30% sample size saving, there would be chances to finish with similar sample size as initially planned., while reducing risk for a negative trial. The time until disease flare: Cox proportional-hazards regression model. Kaplan–Meier estimates plotted against time. Changes in inflammatory markers: Wilcoxon rank-sum test. Secondary variables at other times will be analysed using Fisher's exact tests or ANCOVA as appropriate.
Missing data handling: Patients who met the criteria for relapse or who discontinued treatment prematurely for any reason will be considered to have had a disease relapse for ITT.
Tolerability and acceptability: Fisher's exact test to compare the incidence of infections between study groups

6.2.4. Supportive safety data

Same as main development

7. Analysis of the practical, ethical and regulatory impact

7.1. Option 1: Parallel delayed start pivotal study

Method assessed: Option 1: delayed start randomization	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements	No	Not necessarily, as it requires a concurrent placebo arm in phase 1, thus may need an increment of sample size.
• May shorten time to study completion	No	Larger sample size will take longer to complete recruitment
• May ease recruitment	Yes	All patients will receive active treatment sooner or later no much negative impact expected.
Statistical assessment:		
• Improves internal validity	Yes	Will reduce the placebo effect and any potential effect of the patients knowing he is receiving an active intervention. Therefore, a better estimation of the effect is anticipated. Changes from the baseline measurement within each patient will be used as an outcome measure.
• Increases stability of estimates	No	No changes expected.
• Increases sensibility to changes	No	No changes expected as the study will use the same endpoints

Method assessed: Option 1: delayed start randomization	Improves?	Comments
• Compliant with predetermination	Yes	All parameters and analysis defined a priori
• Consistency (discuss)	No	Less robustness of conclusions. As per secondary variables, larger size may allow detection of differences in secondary vars.
• Robustness of method (discuss)	Yes	A concurrent placebo-control arm will add robustness to the estimation of the treatment effect, with less risk of overestimation of effect than enriched methods.
• Protection against type I and II errors (discuss)	Yes	Similar
Regulatory assessment:		
• Risk of bias and credibility	No, actually may improve credibility	The initial stage, i.e. the parallel double blind design may increase the credibility and reduce the risk of bias compared to the actual design (i.e. less risk to overestimation of the treatment effect)
• External validity (discuss)	Yes	Is not based on enriched population as the actual pivotal trial
• Therapeutic positioning and comparisons	No	Indirect comparison with other placebo controlled trials through study placebo response
• Informative on relevance and clinical impact	Yes	Improved as RCT vs placebo may allow a better estimation of the effect and interpretation of its relevance
• Enough information on safety	No, worsens	Less patients exposed to the relevant dose of active treatment
• Suitable information for risk-benefit balance	Yes	Better estimation of effect size, but smaller population exposed to the relevant dose may slightly reduce safety data in an already limited safety database.
Ethical assessment:		
• May minimise risks	No	Well managed risks of failure, but time on placebo is longer than the actual development
• May maximize access to treatment	Yes	Access to the active treatment is guaranteed by participating in the study
• May minimise unnecessary exposure to ineffective treatments or placebo	No	Some (half of the) patients will receive placebo, although they will receive with some delay the active treatment (in stage II).

Method assessed: Option 1: delayed start randomization	Improves?	Comments
• Considers patient input	No	No improvement, same as before

In summary, option 1 can improve the quality, robustness of the evidence and protection against errors, because it may reduce the placebo effect and any potential effect of the patients knowing that the treatment received is an active intervention. Therefore, a better estimation of the effect is anticipated. On the other hand, this alternative option will not improve practical and ethical aspects, in fact, if any, it would have a negative effect as the addition of a placebo arm may delay recruitment and increase exposure to an ineffective treatment. This alternative may be a suitable option and might have facilitated the regulatory access, but it is noted that the main uncertainties would not be addressed. In less severe conditions, such problems are less an issue.

7.2. Option 2: Multi-arm group sequential design

Method assessed: Option 2: multi-arm multi-stage sequential design with a simultaneous stopping rule	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements	No	Increased because more arms are used
• May shorten time to study completion	Yes	Yes, theoretically.
• May ease recruitment	No	Larger size is more difficult to recruit, and a placebo treatment arm may complicate recruitment.
Statistical assessment:		
• Improves internal validity	Yes	If a dose-response effect is demonstrated, adds to superiority to placebo as an evidence of efficacy
• Increases stability of estimates	No	Probably not, in fact the decision based on the interim analysis may be subject to less stability in the estimates
• Increases sensibility to changes	Yes	By including a placebo arm concurrently
• Compliant with predetermination	Yes	Yes, interim pre-planned and methods for adjustment in place
• Consistency (discuss)	No	The ability to show an effect in secondary endpoints may be reduced if a reduced sample size or shorter term FU
• Robustness of method (discuss)	Yes	Based on pre-established criteria for dose-selection

Method assessed: Option 2: multi-arm multi-stage sequential design with a simultaneous stopping rule	Improves?	Comments
• Protection against type I and II errors (discuss)	Similar	Methods for adjustment in place
Regulatory assessment:		
• Risk of bias and credibility	No	Parallel double blind design, appropriate methods for analysis
• External validity (discuss)	No	Larger size, but small numbers per group, yet inference might lead to few supportive data if early termination.
• Therapeutic positioning and comparisons	No	Compared only with placebo
• Informative on relevance and clinical impact	No	Improved as RCT vs placebo may allow a better estimation of the effect and interpretation of its relevance
• Enough information on safety	Yes	May be reduced, at least for some treatment arms and thus, the overall safety database might be affected
• Suitable information for risk-benefit balance	Yes	Better information on dose selection, dose-response as an additional evidence of efficacy
Ethical assessment:		
• May minimise risks	No	No impact
• May maximise access to treatment	No	Randomisation rate to active:placebo 2:1, but half of active patients assigned to a dose different form the one that will be selected
• May minimise unnecessary exposure to ineffective treatments or placebo	No	It increases, by including placebo and an alternative dose
• Considers patient input	No	Similar to other designs, included as secondary variables

In summary, option 2 may improve credibility in dose-selection, theoretically minimise exposure to ineffective treatments, and reduce time to completion, but given the little room for improvement (due to the overall limited trial size) this are not deemed a major advantage. However, this could address one of the main uncertainties identified in the development plan conducted, i.e. selection of the optimal dose.

8. Recommendations

The development of Ilaris in the treatment of CAPS including MWS, NOMID/CINCA, FCAS/FCU is considered a representative model within the acute repeated episodes cluster of conditions.

Therefore, previous considerations on applicability of novel methods can reasonably be considered suitable options for conditions belonging to the cluster of acute repeated episodes conditions.

In particular, challenge-dechallenge-rechallenge are particularly compelling and convincing to regulators, since causality conclusions derived from temporal sequence of events are robust and intuitive. Besides, from ethical perspective seem to be a good solution in serious conditions when a reasonable plausibility exists. This has been the method applied in the Ilaris pivotal clinical development. The main drawback is that the size of effect is overestimated, since only responders are followed – in that sense, conventional parallel designs may be regarded as less biased.

Sequential designs based on binary assessment of occurrence of first flare may be an efficient approach to these developments. Arm-dropping adaptions may be useful if coupled to sequential approaches to support dose-finding through futility interim assessments – Bayesian methods to discard minimum required activity to continue an arm are good options for implementation of such adaptions.

External controls, if robust, could provide support to single arm trials, but only if the natural history of the condition is invariably showing poor prognosis and the size of effect in treated subjects is outstanding. Rates of flares may be compared between groups, or patients be paired with similar patients in external group, but frequency and type of assessments should be standard enough as to ensure comparability is not impaired.

Longitudinal designs with repeated measures may be applied if the number of events in a period is more important than the time to first episode. In Ilaris example, time free of flares is important, but the number of flares in a period is also a relevant outcome. Designs where repeated measures are taken into account, with or without weighing by severity of flare, could be an alternative for drugs with lower efficacy.

Delayed start is a suitable approach in this setting, since allows to cover ethical concerns (all patients will access the experimental treatment), provides a true baseline assessment, may thus control for overestimation of effect size, and may allow to describe prognostic impact and persistence of effect on the long term.

Classic designs remain a gold standard in terms of robustness, although may be logistically and ethically difficult to implement, and requires higher sample sizes. Use of placebo generally limited in time for non life-threatening conditions or when lack of prognostic consequences for periods without treatment. Rescue plan is needed (either for placebo or for experimental treatment)

Annex 4. 3. Revestive® (teduglutide)

Clinical Development Plan Revestive® (teduglutide) Clinical development in Short Bowel Syndrome

Version 1; 24/098/2017

1 Index

2 Introduction

2.1 Background

2.1.1 Disease and currently available alternatives

Short bowel syndrome (SBS) is a serious, disabling, socially incapacitating and potentially life-threatening condition resulting from surgical resection, congenital defect, or disease-associated loss of intestinal absorption and is characterised by the inability to maintain protein-energy, fluid, electrolyte, or micronutrient balances when on a conventionally accepted, normal diet.

SBS is characterised by large heterogeneity explained by differences in remnant bowel anatomy; some patients are able to compensate for their malabsorption by increasing oral intake and adapt metabolically, but others depend on parenteral infusions.

Parenteral nutrition (PN) can be life-saving but also may be associated with rare potentially life-threatening complications such as catheter related blood stream infections, central venous thrombosis and even embolism. Also parenteral constituents and chronic dehydration may contribute to PN associated liver and renal disease and even eventually to organ failure.

Symptoms of SBS and the inconveniences and complications in relation to parenteral support may cause potential restrictions in lifestyle and impairment of quality of life.

Prognosis depends on length of residual small intestine, residual underlying disease, absence of the terminal ileum, ileocecal valve and a colon in continuity, nature of the primary disorder, and the degree of intestinal adaptation following resection and/or intestinal injury.

There are no established pharmacological treatments available for the SBS condition, which receives only supportive symptoms related drug care.

2.1.2 Rationale for the development

The hormonal stimulation to augment remnant bowel adaptation has been suggested, with glucagon-like peptide 2 (GLP-2), a peptide which is secreted from the intestinal L-cells following meal ingestion, being as a key factor in this respect.

Teduglutide is a novel recombinant analogue for GLP-2, a natural occurring peptide which is secreted primarily by the lower gastrointestinal tract. A slight variation in the peptide confers to teduglutide a longer (up to 2h) half-life than the natural compound (7 minutes).

Due to the high analogy between the two molecules and based on the activation of the G-protein coupled GLP-2 receptor as an in vitro proof of concept, teduglutide is deemed to have the same effects described for GLP-2. GLP-2 activation is thought to result in the release of several growth factors such as IGF-1, EGF and KGF. Native GLP-2 reduces gastric motility, inhibits gastric acid secretion, increases intestinal blood flow, and enhances the transport, absorption and utilisation of nutrients. In addition, it is involved in regulating maintenance and adaptive growth of the small intestinal mucosa.

In the repeat-dose toxicity studies, exaggerated pharmacological effects of teduglutide including biliary and pancreatic duct hyperplasia were observed in mice and monkeys at clinical relevant exposures. The applicant considered the exaggerated pharmacodynamic effects observed at high dose levels as non-adverse as they were fully/partly reversible and there was no indication of tissue dysfunction.

Based on tumour initiation studies in mice, teduglutide may exert a tumour-promoting effect. Teduglutide does not have a direct growth promoting effect on isolated tumour cells, thus it most likely acts indirectly via induction of release of growth mediators. Moreover, in the long term rat carcinogenicity study, jejunal neoplastic findings were made at plasma exposure levels ≥ 8-fold higher than the one observed in patients administered the recommended daily dose. Therefore, considering that teduglutide and GLP-2 induce intestinotrophic effects, it is not unexpected that intestinal neoplasms may occur following life-long treatment of rats. Hence, it cannot be excluded that teduglutide may have a tumour-promoting effect in patients undergoing long-term treatment.

2.2 Scope of development

The product is currently intended for the treatment of adult patients with Short Bowel Syndrome. Patients should be stable following a period of intestinal adaptation after surgery.

The scope of development is from initial clinical data and decision to proceed to regulatory development to pivotal evidence achieved, and consistent with the Target Product Profile

2.2.1 Target product profile

Indication	Treatment of adult patients with Short Bowel Syndrome. Patients should be stable following a period of intestinal adaptation after surgery. (treatment of paediatric patients with Short Bowel Syndrome)
Route of administration	Subcutaneous
Pharmaceutical form	Teduglutide is provided as powder and solvent for subcutaneous injection in strength of 5 mg. This is reconstituted with 0.5 ml of sterilised water for injection, i.e. after reconstitution a nominal 10 mg/ml solution is obtained.
Posology	Daily dose of 0.05 mg/kg body
Main target population	Adult (and paediatric) patients with Short Bowel Syndrome
Claims to be supported by the clinical development.	Improvement in absorption of nutrients and reduction in parenteral nutrition requirements in patients with short bowel syndrome.
Regions where the product should be marketed:	Global (or EU + USA)
Regulatory agencies that will be involved	EMA FDA

3 General investigational plan

3.1 Objective (s) of the development

The objective of the development is to provide pivotal support to the application for marketing authorisation in the EU of the product, by generating:

- Confirmatory evidence of superiority to placebo in short bowel syndrome

- An appropriate safety database, including enough information to permit a characterisation of the safety profile of the product, and a risk-benefit assessment of the product at the time of assessment of the marketing authorisation application in the EU
- Any needed supporting information on the product to apply for a new drug authorisation in EU.

4 Assessment of applicability of methods

4.1 Representativeness of Revestive within the cluster

The development of Revestive in the treatment of Short Bowel Syndrome is considered a representative model within the cluster of chronic stable conditions, since patients have a long lasting condition that has reached a reasonably stable status. However, heterogeneity across potential patients may be more important in this particular development than in other medical conditions within the cluster. This may also explain why the clinical program used strict inclusion and exclusion criteria. As a consequence, the duration of the trials was longer than expected for a prevalent condition, allowing in this way to consider methods using information as acquired – such methods may not be relevant in other medical conditions within the cluster.

4.2 Applicability of novel methodologies based on UMCU report

Applicable methodologies

Long short term outcomes: intermediate variables available with clinical value that may relate to long end-points.

Sample size reassessment and hypothesis testing in adaptive survival trials: Not a survival adaptive trial but key secondary endpoints are time-to-event endpoints. Increased precision for sample size reassessment. Better use of available patients and avoid usage of placebo arm. Logistically increased complexity for trial conduct, management and coordination. More time needed for reassessment.

Multi-arm group sequential designs with a simultaneous stopping rule: Partially applied. Two treatment arms in one pivotal study the result of which was termed hypothesis generating because they used a stepwise procedure with step 1 not significant. Advantages include more compelling/robust evidence regarding the best option for treatment or treatment regimen or the right dose and/or posology. Avoid the use of placebo arm and this would be ethically more confident. Better use of available patients as the total exposure is 190 and 170 randomised in the two placebo-controlled pivotal trials. Disadvantages include: Logistically increased complexity for trial conduct, management and coordination. Interim extra work such as increased readiness required in order to perform the interim analyses as the outcomes timing depends and is closely correlated with the recruitment pattern. Even if the interim analyses are performed earlier than by conducting a traditional trial, this could lead to extra time necessary for instance for an IDMC/DSMB to conclude on the go/ no-go decision and this furthermore could lead to the unfortunate scenario for patients to not be enrolled in the study

until the decision is made. Overall higher economical and administrative burden on trialists and sponsors.

Sequential designs for small samples: Short time to outcome, longer recruitment period, good ratio (recruitment period /time to outcome Years: months (>1)). Applicable if only 1 experimental arm. Advantages include: Fewer patients included in the trial that would lead to fewer centers, which could lead to less time needed for the NECs to approve the study protocol in each country, as well as increased safety for patients. Better use of available patients. Efficacy results obtained earlier would lead to less approval procedural time, hence timely availability of OMP on the EU/EAA market. Optimised use of accumulated knowledge from previous studies reported in literature and increase in precision for the informed prior for treatment effect size estimates that would lead to increase in precision for the adjustment of boundaries depending on the set power. Disadvantages include: Interim extra work, sufficient level of evidence, but not overwhelming (regardless the positive or detrimental effect on patients); however, patients could be followed prospectively (as part of a registry).

Bayesian sample size re-estimation using power priors: Applicable if a continuous key secondary endpoint and one experimental treatment. Second study. Advantages are fewer patients. Disadvantages include possible extra patients in case of effect size overestimation.

Dynamic borrowing through empirical power priors that control type I error: Applicable if prior information on effect size and one experimental arm only. Second study. Advantages are fewer patients. Disadvantages include possible extra patients in case of effect size overestimation.

Minimisation and stratification: Already applied, useful to manage heterogeneity and to ensure pre-determined subgroup analysis.

Might be applicable

Fallback tests for co-primary endpoints: if multiple relevant end-points without clear prioritization of a lead main variable.

Optimal exact tests for multiple binary endpoints: if multiple relevant end-points without clear prioritisation of a lead main variable.

Goal Attainment Scale: Could be applicable to cluster, but not relevant in this case according to UMCU.

Not applicable

Heterogeneity estimators

Prior distributions for variance parameters in sparse-event MA

Simultaneous inference for multiple marginal GEE models

Additional considerations based on disease clustering

Intrasubject comparisons: Because of the relatively stable course of the condition, alternative designs modifying group setting may be applicable, such as intrasubject comparisons or start-stop designs.

Goal Attainment Scale: The difficulties for interpretation of clinical relevance and heterogeneity in clinical presentation support that a Goal Attainment Scale may be used as secondary variable to provide support for clinical relevance of findings, especially if methods that allow concluding on small samples, are applied.

Stratification: due to heterogeneity, stratification by severity is required to manage analysis of subgroups (already applied); in this case, GAS is even more useful to guide regulatory assessment of clinical relevance.

Methods using information as acquired: Because of the relatively stable course of the condition, patients are supposed to be prevalent rather than incident, so that the duration of the trial should be more guided by logistic setting-up and patient follow-up than for identification of subjects and recruitment, and adaptions and sequential approaches would likely provide little advantage on other methods. However, both pivotal trials had long recruitment periods (19 and 33 months, respectively) so that in this case these methods could have been applicable.

5 Actual development plan for Revestive

Revestive has been studied after subcutaneous and intravenous injection in seven phase 1 study. Subcutaneous administration has been used for the phase 2 and for two phase 3, double blind placebo controlled studies. Long term safety of Revestive has been further evaluated in two open label studies, extension of the two phase 3 studies.

5.1 Safety and tolerability

- Single ascending dose safety and tolerability study of 4 doses and placebo in 32 healthy male subjects.
- Multiple ascending dose safety and tolerability study of 7 doses given once or twice daily subcutaneously in the abdomen to 95 healthy male and female subjects for 8 days.
- Healthy subjects PD and PK/PD C09-001.
- Study of the effect of a single dose of Teduglutide on cardiac repolarisation (QT, QTc interval).

5.2 Pharmacokinetics

- PK bioavailability open label two ways crossover study comparing iv infusion and sc administration in 14 healthy male and female subjects.
- PK bioavailability open label three ways crossover study comparing bioavailability of an SC injection of 10 mg Teduglutide in the thigh and arm, relative to the abdomen in 18 healthy male and female subjects.

- Open label study in moderate hepatic impairment on Teduglutide PK following SC administration of a single 20 mg dose to 12 hepatically impaired subjects as compared to 12 healthy matched controls.
- Open label study in renal impairment on Teduglutide PK following SC administration of a single 10 mg dose to 36 patients with moderate or severe renal impairment, or end stage renal disease, or healthy subjects.

5.3 Proof of activity/dose finding

- Phase 2, open-label, multicentre, dose-ranging, pilot study to study safety and tolerability of a 21-day, ascending, multidose, SC administration of Teduglutide (ALX-0600) in SBS patients, PD and PK/PD of 3 doses (0.03, 0.1, 0.15 mg/kg/ day) administered SC once (qd) or twice daily (bid) for 21 days to 17 male or female SBS patients without colon or with ≥50° of their colon continuity.

5.4 Dose response studies

Study 92001: Open-label-multicenter, dose-ranging, pilot study to examine the safety, tolerability and effect of a 21 day, ascending, multidose subcutaneous treatment with ALX-0600 (Teduglutide) in patients with short bowel syndrome.

5.4.1 Dose finding study

The study was a proof of concept dose exploration that failed due to poor recruitment and lack of appropriate standardisation methods for measurement of main pharmacodynamics assessment, raising inconclusive results.

Title of study: Open-label-multicenter, dose-ranging, pilot study to examine the safety, tolerability and effect of a 21 day, ascending, multidose subcutaneous treatment with ALX-0600 (Teduglutide) in patients with short bowel syndrome (Study 92001)	
Investigators (Study center): 5 trial centers in the US and Europe (Denmark)	
Studied period: 1-2 years First patient included on 2000 Last patient completed on 2001	**Phase of development:** Phase II
Objectives **Primary:** to determine the safety and tolerability of a 21-day s.c. dosing regimen of teduglutide in SBS patients. **Secondary:** to measure the PD effect on the capacity of the remaining bowel to absorb water and macronutrients after 21 days treatment.	
Design: Phase 2, open-label, multicentre, dose-ranging, pilot study. The study consisted of four phases: screening phase (Days -14 to -1), a pre-dose phase (Days -3 to -1), 21-day dosing phase, and 21-day follow-up phase. Eligible patients were admitted as inpatients to hospital wards or General Clinic Research Centres (GCRC) on three separate occasions, 18 days apart, for the last four days and three nights of the baseline period and at the end of the treatment and follow up periods.	
Number of patients, by arm: 17, in 5 groups of 3, 5, 4, 5 and 5 patients, respectively.	

Intended sample size: not disclosed **Populations for analysis:** Only those compliant with the protocol.
Main inclusion criteria: males and females over 18 years of age with a diagnosis of SBS due to vascular ischemic disease, malrotation, or volvulus, or with quiescent IBD. Patients should have undergone intestinal resection at least 12 months before entering the study and residual small intestine had to be less than 150 cm in length. Patients had also to present with normal body weight, normal albumin levels, and no use of glutamine supplementation for at least 4 weeks. Patients without colon were eligible and had to have at least 50% of their caloric intake by PN; patients with colon had to have at least 50% of their colon preserved with a fecal weight exceeding 1.0 kg/day (of a 72 hour stool collection period) and fecal energy loss of more than 2.0 MJ/day (or fecal fat loss of more than 50 g/day on their habitual diet).
Main Exclusion criteria: Patients with active IBD, history of pseudo obstruction, recent surgery, fistulae, and others had to be excluded.
Test product, dose and mode of administration, duration: Single doses up to 10 mg (equivalent to approximately 0.10 mg/kg) were shown to be safe and well tolerated when administered to normal volunteers (previous study, data not presented). Based on the pharmacological, toxicological, and pharmacokinetic profile available when the protocol was prepared, doses of teduglutide 25% log above (0.15 mg/kg) and 50% log below (0.03 mg/kg) the 0.10 mg/kg dose were included in this pilot study. Five patients from the two highest dosing groups were rechallenged with a divided dose (0.05 mg/kg or 0.075 mg/kg every 12 hours) at least three months after their initial participation to investigate antibody formation and to evaluate twice daily dosing. Five SBS patients with ⩾50% colon in continuity received teduglutide 0.10 mg/kg/day once daily. 0.03 mg/kg/d qd x 21 days (N=3) 0.1 mg/kg/d qd x 21 days (N=5) 0.15 mg/kg/d qd x 21 days (N=4) 0.05 mg or 0.075 mg/kg/d bid x 21 days (N=5) 0.1 mg/kg/d qd x 21 days (N=5 - colon)
Reference therapy, dose and mode of administration, duration: **Active:** None; **Placebo:** None
Criteria for evaluation (Definition, timing of assessments): **Pharmacokinetic data / Pharmacodynamic data / other:** For the purpose of the PD/clinical parameters evaluation, three 72 hour nutrient absorption tests were used to measure the capacity of the intestine to absorb macronutrients. **Main efficacy assessment:** Not applicable

Secondary variables: Urinanalysis test were used both for safety and for determining fluid excretion via the kidneys. D-Xylose tests with plasma and urine collections were also performed as a measure of malabsorption.
Statistical methods **Analysis of efficacy**: A test for normality was performed justifying parametric statistical testing in this study. A Student's paired *t* test was employed that compared treatment with baseline values, and follow up with baseline values. Absolute intestinal absorption was calculated as equivalent to the difference between ingestion and excretion, and relative absorption as (diet−faecal)/ diet×100%. No comparisons were made between patients on different doses or with different anatomy due to the limited number of patients in the study. Data are expressed as mean (SD). A value of p<0.05 was considered significant. **Missing data handling:** NA **Tolerability and acceptability:** Descriptive
Main results (only actual scenario): The study was closed before all planned patients had been enrolled due to slow enrolment. Four US sites recruited six (two, two, one, one) patients and one Danish site recruited 12 patients, in total 18 patients were enrolled, 17 dosed, of which 2 were discontinued prematurely. The clinical evaluation part of the study was conducted on the PP population, which consisted of 16 patients (1 patient was excluded due to febrile episodes considered unrelated to study drug). The demographic characteristics of the study groups were approximately similar. Of the patients, 8 (47%) were male, 16 (94%) were Caucasian, and the mean age was 49 years, with a mean weight of 59 kg and mean height of 168 cm. 12 of the 17 patients had Crohn's Disease as underlying cause of SBS. The results from the 72-h nutrient absorption test were not consistent. Variations in the quantity and composition of ingested food between study periods occurred. Pooled groups were analysed and did reveal more consistent results.

5.5 Pivotal evidence (2 studies)

- **CL0600-004:** 24-week double-blind, randomised, placebo-controlled, parallel group study comparing the efficacy, safety and tolerability of two doses of teduglutide (0.05 mg/kg/day and 0.10 mg/kg/day) and placebo in subjects with Parenteral Nutrition-Dependent Short Bowel Syndrome.
- **CL0600-020:** A 24-Week Study of the Efficacy and Safety of Teduglutide in Subjects with Parenteral Nutrition-Dependent Short Bowel Syndrome. A randomised, double-blind, placebo-controlled, parallel-group study.

5.5.1 Pivotal study (1)

Title of study: Study CL0600-004: 24-week double-blind, randomised, placebo-controlled, parallel group study comparing the efficacy, safety and tolerability of two doses of teduglutide (0.05 mg/kg/day and 0.10 mg/kg/day) and placebo in subjects with Parenteral Nutrition-Dependent Short Bowel Syndrome	
Investigators (Study center): 32 centers in the USA (15), Canada (4), and European countries (13 centers in the UK, France, Denmark, Poland, the Netherlands, Belgium, and Germany).	
Studied period:	**Phase of development:**

First patient included on 25th May, 2004, and the last subjects' evaluations were performed on 06th July, 2007. The recruitment lasted 33 months + 6 months follow-up of last randomised patient.	Phase III
Objectives **Primary:** to evaluate the efficacy, safety, tolerability and pharmacokinetics of teduglutide compared with placebo in subjects with parenteral nutrition (PN)-dependent Short Bowel Syndrome (SBS).	
Design: Randomised, double-blind, placebo-controlled, parallel-group multicentre study that aimed at evaluating the efficacy, safety, tolerability and pharmacokinetics of teduglutide in patients with PN-dependent SBS. Run-in with optimisation of PN (3 days to 8 weeks) and stabilisation of PN (4-8 weeks); then stratified randomisation (by ≤6 or >6 L/week of PN at baseline) to parallel comparison of three groups (low dose, high dose and placebo) for 24 weeks; then continued in a separate extension study where patients in placebo were crossed to active (randomisation to low or high) and patients in active continued with same treatment for 28 additional weeks. Randomisation was balanced for treatment groups, participation in the 72-hour nutrient absorption test, and PN at three levels of consumption [PN consisting of IV fluid and electrolytes only (3-7 times weekly), PN 3-5 times weekly, and PN 6-7 times weekly]. Subjects were randomised across centers rather than within a center. The study was double blind, but swelling of stoma is as an AE to teduglutide that may break the blind. Persons responsible for assessment of PN requirements and adjusting doses of PN were to be different from the ones conducting physical examinations and assessing safety.	
Number of patients, by arm: 80 in total, 32 subjects in each of the teduglutide treatment groups and 16 subjects in the placebo group. **Intended sample size:** To detect an anticipated minimum response (20% decrease in PN for Weeks 20 to 24) in 5% of the placebo treated- and 50% of the teduglutide patients with 90% power, 80 subjects were planned to be randomised in a 1:2:2 ratio to one of three treatment arms: placebo, 0.05 mg/kg/day teduglutide, or 0.10 mg/kg/day teduglutide. **Populations for analysis:** ITT included all subjects randomised was the primary study population.	
Main inclusion criteria: SBS due to the most common causes of SBS (surgical resections due to Crohn´s disease, cancer, vascular insufficiency and volvulus). Patients had to be dependent on PN at least 3 times weekly for at least 12 months. Prior to randomisation to study treatment patients had to be stable in their disease for the past 4 weeks as measured by usage/volume of PN, urinary output, urine sodium, renal function, haematocrit, and motility altering medications	
Main Exclusion criteria: SBS patients with possible fluctuating activity in the disease (due to radiation therapy, active Crohn´s disease and celiac disease), history of cancer or clinically significant lymphoproliferative disease, history of alcohol or drug abuse (within previous year) and failure to washout of forbidden medications.	

Test product, dose and mode of administration, duration: Teduglutide (0.05 or 0.10 mg/kg/day) was administered by the s.c. route once daily into one of the four quadrants of the abdomen or either thigh, for 24 weeks.

Reference therapy, dose and mode of administration, duration:
Active: none
Placebo: matching placebo for 24 weeks; teduglutide and placebo vials and their content were identical in appearance.

Criteria for evaluation (Definition, timing of assessments):
Efficacy assessments:
Timing of assessments: Weeks 0, 4, 8, 12, 16, 20, and 24 (main).
Main efficacy assessment: N (%) of subjects who demonstrated a response (≥20% reduction from Baseline in weekly PN volume) at Week 20 maintained at Week 24. The main end-point was amended later to ordered categorical (graded) criterion accounting for duration (weeks 16 and 20 and weeks 20 and 24) as well as intensity of response (percentage reduction in PN volume; 20-100%) meaning that subjects with larger, earlier and/or more sustained response had higher "weight" in the outcome (score of 0-5).
Secondary variables: N(%) of subjects who demonstrated a response (at least a 20% reduction from baseline in the weekly PN volume) at week 20, and who maintained that response at week 24; N(%) of subjects who achieved at least a 1-day reduction in weekly PN; the absolute reduction from baseline in weekly PN kilojoules; absolute reduction of weekly volume of PN from baseline; change from baseline in plasma citrulline at dosing week 24. Exploratory: Time to 20% reduction in PN volume, time to discontinuation of PN, time to a 1-day reduction in weekly PN, N(%) of subjects with reduced i.v. catheter access at week 24, change from baseline in bone markers BSAP and NTx, in lumbar spine ad hip BMD, and in PTH at week 24, change from baseline in QoL at weeks 4, 8, 12, 16, 20, and 24 (scores to be used were: SF-36, IBDQ, and abbreviated EuroQol)., the mucosal crypt-villus architecture and cellular composition within the small and large intestine.

Safety: Laboratory safety samples were evaluated at least once during the seven days following a PN reduction, accompanied by determination of 48-hour urine output. In addition, evidence of dehydration, such as body weight, clinical signs and symptoms, was assessed day 3 to 4 and day 6-7 following a reduction.
Pharmacokinetic data / Pharmacodynamic data / other:

Statistical methods
Analysis of efficacy: N (%) of responders (20 to100% reduction in PN/i.v. volume) were compared by Cochran-Mantel-Haenszel (CMH) test statistics adjusted for the randomisation stratification variable (≤6 or >6 L/week of PN at baseline). The percent and absolute change in PN/i.v. volume from baseline to the last dosing visit, as well as all scheduled visits starting at Week 4 were compared using an analysis of covariance (ANCOVA) model with effects for treatment and baseline PN volume, with the potential for the interaction of the two variables also included as an effect. The least squares means and standard error, along with 95% CIs, were presented for each treatment. Duration of response based on the number of consecutive visits (categorized as 0, 1, 2, and ≥3) for which the subject had a 20 to 100% reduction in weekly PN/i.v. volume from baseline at Week 24 plus previous scheduled visits with a 20 to 100% reduction from baseline, and N (%) of subjects with each level of the

graded (or ordered categorical) response, were compared using extended CMH test statistics (with standardized mid-ranks) adjusted for the randomisation stratification variable.
Missing data handling: There was a requirement to have at least 9 out of 14 daily PN volume data points for the weekly PN volume calculation. If less data were available for the specific period of time, the patient was left out of the analysis for that specific time point. Thus, an imputation method for missing data for the weekly PN volume calculations was not implemented. The applicant subsequently provided evidence that the three study groups were well balanced as regards overall frequency of patients with missing values and frequency of patients with "many missing values" (i.e. more than ten). It was concluded that differential "missingness" is unlike to have had any major effect on the results of the study.

Main results (only actual scenario)
No statistical significant difference between teduglutide 0.10 mg/kg/day vs. placebo in the graded response score (p=0.161) used as primary endpoint. There were 1 (6.3%), 16 (45.7%) and 8 (25%) responders in the placebo, 0.05 mg/kg/day and 0.1mg/kg/day group, respectively. The total effect of teduglutide (pooled response: 35.7% were responders) was much lower than expected in the power calculation, thus forming the basis for a type 2 error. No effect on quality of life was detected.

There was no statistical significant difference between teduglutide 0.10 mg/kg/day vs. placebo in the graded response score (p=0.161) used as primary endpoint. According to the statistical analysis plan (SAP) no further statistical testing was then planned. Still, a decision to proceed to analysis of the low dose was taken based on clinical evidence of efficacy.

The post-hoc analysis showed significant differences for teduglutide 0.05 mg/kg/day vs. placebo (p=0.007). There were 1 (6.3%), 16 (45.7%) and 8 (25%) responders in the placebo-, 0.05 mg/kg/day- and 0.1mg/kg/day group, respectively. The total effect of teduglutide (pooled response: 35.7% were responders) was much lower than expected in the power calculation, thus forming the basis for a type 2 error. Most responders had a score of 1 and 2 (of note one placebo subject – the only responder- had a score of 2). Still, 19 patients (54%) and 24 patients (75%) in the low- and high dose groups were non-responders with a score of 0 (<20% reduction in weekly PN volume).

Overall, these results obtained in the low dose group are promising but can only be considered for hypothesis generation.

No effect on quality of life was detected by use of the SF-36 health survey, EuroQoL EQ-5D and the Inflammatory Bowel Disease Questionnaire (IBDQ) after 24 weeks treatment with teduglutide. It was acknowledged that these questionnaires have not been developed for assessment of QoL in patients with SBS, and that disease specific questionnaires were not available, when the study was conducted. Considering the low numbers of patients included in each treatment group in addition to the heterogeneity in symptoms in between SBS patients, it is conceded that these tools may not have been appropriately sensitive to catch any potential difference.

As study CL0600-004 was not a fully satisfying study, unable to provide confirmatory evidence of efficacy, the applicant subsequently performed a new confirmatory study, CL0600-020, testing the dose of teduglutide which in the exploratory analysis of study CL0600-004 was found to be effective.

5.5.2 Pivotal study (2)

Title of study: Study CL0600-020: A 24-Week Study of the Efficacy and Safety of Teduglutide in Subjects with Parenteral Nutrition-Dependent Short Bowel Syndrome. A randomized, double-blind, placebo-controlled, parallel-group study.	
Investigators (Study center): 27 sites in 10 countries (USA 6, Canada 4, Poland 4, Germany 2, Italy 3, France 2, Spain 2, United Kingdom 2, Denmark 1, Netherlands 1).	
Studied period: The first subject was enrolled on 25 November 2008, and the last subject, last visit was on 04 January 2011. The recruitment lasted 21 months + 6 months follow-up for last patient.	**Phase of development:** Phase III
Objectives **Primary:** The objectives of this clinical study were to evaluate the efficacy, safety, and tolerability of teduglutide compared with placebo in SBS subjects dependent on parenteral support (PN and/or i.v. fluids). **Secondary:** Subject's qualify of life (QoL) was evaluated by using a newly developed subject-reported outcome SBS-specific QoL questionnaire.	
Design: Double blind randomised parallel groups comparing one dose of tedugutide with placebo (1:1 ratio), stratified randomisation by 2 levels of baseline PN/i.v. volume (≤6 L/week or >6 L/week), randomised across centers rather than within centers. Randomised, double-blind, 2-arm, placebo-controlled, parallel-group, multinational, multicenter, 2-stage study. Subjects were randomised to either teduglutide or placebo (1:1). Stage 1 included a screening visit; an optimisation period, if needed, of a maximum of 8 weeks, and a stabilisation period that demonstrated stable administration of parenteral nutrition/intravenous (PN/I.V.) volume for a minimum of 4 weeks up to a maximum of 8 weeks. Stage 2 lasted 24 weeks. An extension study followed for 2 years in the scope of trial CL0600-021.	
Number of patients, by arm: **Intended sample size:** 86 randomised at a 1:1 ratio, to detect a difference in responder rates between teduglutide and placebo groups of 35% and 6%, respectively, α alpha=0.05, 2-sided test and power=90%, based on a Fisher's Exact Test. **Populations for analysis:** ITT included all subjects randomised was the primary study population, (with per protocol analysis supportive), based on data from scheduled visits at Weeks 20 and 24.	
Main inclusion criteria: SBS due to the most common causes of SBS (surgical resections due to Crohn´s disease, cancer, vascular insufficiency and volvulus). Patients had to be dependent on PN at least 3 times weekly for at least 12 months. Prior to randomisation to study treatment patients had to be stable in their disease for the past 4 weeks as measured by usage/volume of PN, urinary output, urine sodium, renal function, haematocrit, and motility altering medications	
Main Exclusion criteria: SBS patients with possible fluctuating activity in the disease (due to radiation therapy, active Crohn´s disease and celiac disease), history of cancer or clinically significant lymphoproliferative disease, history of alcohol or drug abuse (within previous year) and failure to washout of forbidden medications.	
Test product, dose and mode of administration, duration: Teduglutide (0.05 /kg/day) was administered by the s.c. route once daily for 24 weeks.	
Reference therapy, dose and mode of administration, duration: **Active:** none **Placebo:** Teduglutide and placebo were identical in appearance. The study center personnel, the sponsor, and all personnel associated with the monitoring or data management for the clinical study were blinded to the treatment assignment. However since the test drug may induce visible changes in the appearance of the intestinal mucosa, patients with jejunostomy/ileostomy may be aware of what kind of the treatment they receive.	
Criteria for evaluation (Definition, timing of assessments): **Efficacy assessments:** **Timing of assessments:** Weeks 0, 4, 8, 12, 16, 20, and 24 (main). **Main efficacy assessment:** N(%) of subjects who demonstrated a response (≥20% reduction from Baseline in weekly PN volume) at Week 20 maintained at Week 24. **Secondary variables:** Reductions in PN/i.v. volume or the direct effects of improved intestinal absorption of fluid; duration of response (ie., total number of weeks at ≥ 20% reduction from Baseline); the proportion of subjects with a ≥ 20% reduction or a ≥ 2 liter (L) reduction from Baseline in weekly PN at Week 20 and maintained through Week 24; N(%) of subjects who stopped PN and time of discontinuation; and absolute change and percent change in PN between Baseline and last dosing visit. Ordered categorical (or graded) response that accounts for both intensity and duration of the response at the end of the 24-week treatment period. Subjects' quality of life (QoL) using a subject-reported outcome SBS specific QoL scale. **Safety:** Adverse events, 12-lead ECG, vital signs, laboratory safety data, antibodies to teduglutide and/or E. coli protein (ECP), and changes in urine output and body weight.	
Statistical methods **Analysis of efficacy:** Primary endpoint responder rates were evaluated using CMH test statistics adjusted for the randomisation stratification variable (<=6 or >6 L/week of PN at baseline). Exploratory endpoint responder rates were evaluated using logistic regression with effects for treatment group and Baseline PN volume. **Missing data handling:** Sensitivity analyses using different ways of handling missing data, demonstrated that the results were not sensitive to this factor. Patients who discontinued the study prematurely was assigned the score "0" for the response variable (i.e. no response). This is considered the most conservative approach. **Tolerability and acceptability:**	
Main results (only actual scenario)	

Because of strict inclusion/exclusion criteria and diversity of the target population, there was a relative large number of screened patients compared to the relative small number of patients randomised. Compared to placebo, teduglutide had statistically significant effect on the primary efficacy parameter, 20% or greater reduction in volume of PN/i.v. at weeks 20 and 24. Non-responder n (%) 16/43 (37.2) Teduglutide 30/43 (69.8) placebo. Responder n (%) 27/43 (62.8) Teduglutide 13/43 (30.2) placebo (p = 0.002). The results were robust and confirmed in a number of sensitivity analyses. The effect persisted across subgroups. The secondary endpoints generally supported the primary endpoint. Compared to placebo, teduglutide had superior effect after 8 to 12 weeks, and had longer duration of effect than placebo in reducing the need for PN/i.v. in SBS. No statistically significant difference between teduglutide and placebo in terms of quality of life using SBS-QoL™ or complete weaning off PN/i.v. was observed. The clinical relevance of the primary efficacy endpoint was acknowledged by EMA, supported by an ad-hoc expert group.

5.6 Supportive confirmatory efficacy and safety data

- Double-blind randomized, parallel-group, multinational, multicenter 28-week extension study to the study CL0600-004
- Long-term, open-label study extension to the study CL0600-020

5.6.1 Supportive (1)

Title of study: CL0600-005 double-blind randomised, parallel-group, multinational, multicenter 28-week extension study to the study CL0600-004	
Investigators (Study center):	
Studied period: Not available	**Phase of development:** Phase III
Objectives **Primary:** Evaluate the long-term safety and efficacy of teduglutide	
Design: double-blind randomised, parallel-group, multinational, multicenter 28 week extension of previous pivotal study. Patients from the active treatment arms continued treatment. Previous placebo patients were randomised to receive 0.05 or 0.10 mg/kg/day of teduglutide. After completing 28-week dosing, all entered a 4-week non-dosing follow-up period.	
Number of patients, by arm: 65 patients entered the study (25 continued into 0.05 mg, 27 continued into 0.1 mg and 13 were naïve: 6 assigned to 0.05mg and 7 to 0.1 mg) **Intended sample size:** based on patients completing previous study	
Main inclusion criteria: Patients completing the study CL0600-004	
Main Exclusion criteria:	
Test product, dose and mode of administration, duration: Patients from the active treatment arms continued same treatment. Previous placebo patients were randomised to receive 0.05 or 0.10 mg/kg/day of teduglutide.	
Reference therapy, dose and mode of administration, duration: **Active:** No	

Placebo: No
Criteria for evaluation (Definition, timing of assessments): **Timing of assessments:** **Main safety assessment:** **Secondary efficacy variables:** **Pharmacokinetic data / Pharmacodynamic data / other:**
Statistical methods **Analysis of efficacy:** descriptive **Missing data handling:** No data **Tolerability and acceptability:** No data
Main results (only actual scenario) 65 patients entered the study (25 continued into 0.05 mg, 27 continued into 0.1 mg and 13 were naïve: 6 assigned to 0.05mg and 7 to 0.1 mg). 54 subjects completed, one was completely weaned of PN. Sustained response (20% reduction from baseline) was 17/25 (68%), 14/27 (52%) in continued patients. Naïve patients responded 5/6(83.3%) and 3/7(42.9%). There were 10 discontinuations: 9 from continued therapies and 1 from naïve patients. The proportions by dose were 20%, 14.8%, 0% and 14.3%, respectively. 25% of the responders during study 004 became non-responders at the end of study 005. Quality of Life data indicated no systematic improvement in QoL scores.

Study CL0600-005 was a double-blind 28-week extension study to the study CL0600-004, and was designed as a randomised, parallel-group, multinational, multicenter study. This supportive study had the objective to evaluate the long-term safety and efficacy of teduglutide. 65 of the 71 patients who had completed the study CL0600-004 were enrolled. The patients from the active treatment arms maintained their treatment regimen. The previous placebo patients were randomised to receive 0.05 or 0.10 mg/kg/day of teduglutide. The patients who completed the 28-week dosing period of the study CL0600-005 entered a 4-week non-dosing follow-up period.

5/25 (20%) patients in the 0.05/0.05 group and 4/27 (14.8%) of the 0.10/0.10 group discontinued. There were 0/6 and 1/7 discontinuations in the placebo/0.05 and the placebo/0.10group, respectively. 54 of the 71 patients completed study CL0600-005; 25% of the responders during study 004 became non-responders at the end of study 005. Quality of Life data indicated no systematic improvement in QoL scores.

5.6.2 Supportive (2)

Title of study: CL0600-021 long-term, open-label study extension of CL0600-020	
Investigators (Study center):	
Studied period: Study Start Date: 21 September 2009 Study Completion Date: ongoing at time of assessment, interim data cut-off 30th June 2011	**Phase of development:** Phase III
Objectives	

Primary: **Secondary:**
Design: long-term, open-label study extension. Patients in placebo were started on active, patients in active were continued.
Number of patients, by arm: In the previous Trudy there were 86 (43/43) patients in active and placebo. The extension included 88 patients, 37 continuations (teduglutide in previous study) and 51 naïve (placebo in study or new patients) **Intended sample size:** based on patients completing previous study (+ naïve patients) **Populations for analysis:**
Main inclusion criteria: Patients completing the study CL0600-021
Main Exclusion criteria:
Test product, dose and mode of administration, duration:
Reference therapy, dose and mode of administration, duration: **Active:** No **Placebo:** No
Criteria for evaluation (Definition, timing of assessments): **Timing of assessments:** information collected at completed visits in the first 6 months (through Visit 6 of this extension study) **Main safety assessment:** **Secondary efficacy variables:** **Pharmacokinetic data / Pharmacodynamic data / other:**
Statistical methods **Analysis of efficacy:** descriptive **Missing data handling:** No data **Tolerability and acceptability:** No data
Main results (only actual scenario) Out of 43 active patients in double blind study, 34 entered the extension and 31 remained within response (20 to 100% decrease from baseline in weekly PN/i.v.) after 6 months of treatment (72.1%, or 91.2% of those entering the extension). This corresponded to mean reduction of 5.16 L/week from a baseline level of 12.76 L/week (relative reduction of 34.2%). 18/43 (41.9% or 52.9% of those entering the extension) achieved at least a 1 day reduction in the number of days on PN/i.v. For naïve patients, 17/43 (39.5%) had response (20 to 100% decrease from baseline in weekly PN/i.v.) after 6 months, with reductions of 2.19 L/week from a baseline level of 11.97 L/week. 10/43 (23.3%) achieved at least a 1 day reduction in the number of days on PN/i.v. Selection of patients based on previous response led to an overrepresentation of subjects who tolerate teduglutide and had experienced a positive effect of the drug in TED/TED group.

Study CL0600-021 was designed as a long-term, open-label study (ongoing at time of assessment). An interim analysis presented information collected at data cut-off on 30 June 2011. Subjects came from Study CL0600-020 (already exposed for 24 w, TED/TED group,

N=37); 2) and naïve to active because of receiving pcb in previous study (PBO/TED group, N=51). Three subjects were weaned from PN/i.v. (group not reported).

Response during 24 weeks of treatment in Study CL0600-020 was maintained in the TED/TED group; mean reduction was 5.16 L/week from a baseline level of 12.76 L/week, with a mean reduction of 34.18%; PBO/TED Group showed reductions of 2.19 L/week from a baseline level of 11.97 L/week. 31 of 34 subjects (91.2%) in the TED/TED group and 17 of 43 subjects (39.5%) in the NT, PBO/TED group had a 20 to 100% decrease from baseline in weekly PN/i.v. volume. 18/34 (52.9%) and 10/43 (23.3%) achieved at least a 1 day reduction in the number of days on PN/i.v.

Selection of patients based on previous response led to an overrepresentation of subjects who tolerate teduglutide and had experienced a positive effect of the drug in TED/TED group. The groups were not comparable in terms of safety and efficacy, so the magnitude of the effect observed and the clinical relevance of the observed effect have to be interpreted with caution.

5.7 Total patient exposure in the target indication

A total of 59 patients were exposed to placebo during the placebo controlled SBS studies with an overall placebo exposure of 26.2 person years. A total of 109 patients were exposed to teduglutide during the placebo controlled SBS studies; 77 patients to 0.05 mg/kg/day and 32 patients to 0.10 mg/kg/day, with an overall exposure to teduglutide of 46.5 person years.

5.8 Additional information for modelling

The results from the failed phase III pivotal trial comparing 2 doses vs placebo showed that the low- and high dose of teduglutide induced mean (SD) reductions in weekly PN volumes from baseline to week 24 of – 2.48L (2.34) and -2.47L (3.33), respectively. The corresponding reduction for placebo was -0.90L (1.41).
For high dose: parenteral volume was gradually decreased by 159±342, 242±361, 275±367, 309±431 and 353±475 ml/day at weeks 8, 12, 16, 20 and 24, respectively (all $p<0.05$) compared with baseline. These parenteral volume reductions were not significant when compared with placebo ($p=0.08$ at week 24). For low dose: parenteral volume significantly decreased by 109±230, 193±223, 246±277, 319±295 and 354±334 ml/day at weeks 8, 12, 16, 20 and 24 (all $p<0.05$), respectively, compared with baseline. A minor but statistically significant reduction in parenteral volume was observed in patients in the placebo group at weeks 12 and 24 compared with baseline (106 +/-167 ml/day, $p=0.02$ and 128+/-202 ml/day, $p=0.03$, respectively).
Combined analysis of study CL0600-020 and its prolongation study, CL0600-021, show that in the time period up to 6 months after initiating treatment there is a gradual increase in the number of patients achieving the primary endpoint. After that time point only few additional patients achieve the primary endpoint. Thus the proposal to evaluate efficacy after 6 months is supported.

5.9 Uncertainties/weaknesses identified

An intermediate variable is available that is deemed reasonably representative of the clinical goal, but care should be paid to the relaxation of the treatment goal when an intermediate

variable is used, so that changes that are significant in the intermediate variable are then discussed for relevance, and the initial treatment objective is forgotten thereafter.

In three of the 5 studies dedicated to the primary PD, plasma citrulline was measured being regarded as a biomarker of enterocyte mass. In general, plasma levels of citrulline increased after teduglutide treatment. However, the increase in citrulline was not associated with a decrease in PN volume. Hence, a key role of plasma citrulline in assessing the PD effect of teduglutide is not supported, and thus has not been further assessed. Main marker of activity since then has been the PN volume requirements. Volume requirements will be highly dependent on the baseline values, and changes may be substantially different depending on the relative percentage they represent vs the baseline. Heterogeneity in the baseline values for volume requirements is key in this particular medical condition.

The dose selection was based on a very small exploratory trial with only 16 patients that was inconclusive due to heterogeneity and lack of standardised protocols for food intake. This is an additional source of heterogeneity that difficult to obtain a reliable proof of efficacy.

The confirmatory development was based on a first pivotal trial comparing two doses to a placebo, in a 2:2:1 ration, that had a major amendment during the study conduction, in which the main end-point was changed to qualitatively account for heterogeneity. However, the variable failed and the study had a negative conclusion for its main objective. The main end-point was not met for the high dose, but was for the low dose; thus the study had to be repeated and a second pivotal was done, using the low dose only. In the first pivotal study, heterogeneity was substantial and the investigators suggested that baseline values were not well balanced.

The second study concluded significant differences vs placebo in a 1:1 randomised treatment vs placebo. Heterogeneity was lower, but external validity was impaired in exchange. The main end-point was simpler than in the former failed pivotal trial, and statistical significance was met, although not at the size of effect that was anticipated from the mechanism of action. The effect is thus focused on reducing weekly volume of PN requirements, and not being able to avoid PN completely. Such goal has been achieved rarely, often in open-label extensions on the long term, which may be confounded by the natural course of the condition. While the clinical relevance of the primary efficacy endpoint was acknowledged by EMA, they required the support of an ad-hoc expert group to conclude on that.

Risks may include malignancy, again considering the mechanism of action of teduglutide. A clinical development with a perspective of 1 year (or less) is unable to detect such risks, unless their magnitude is outstanding. The risk/benefit assessment, thus, is uncertain in both the relevance of the clinical changes and the magnitude and severity of potential adverse reactions.

6 Alternative development plans

6.1 Option 1: group sequential study with sample size reassessment and simultaneous stopping rule

This option was selected based on the results of the systematic exercise of applicability of methods. A design including three different features of the recommended methods has been proposed, including: sample size reassessment, multi-arm group sequential trial with simultaneous stopping rule for continuous variable, and stratified randomisation. A Goal Attainment Scale will be used as a secondary variable to provide support to the clinical relevance of any observed change.

6.1.1 Safety and tolerability

Same as actual development

6.1.2 Proof of activity/ dose exploration

Same as actual development

6.1.3 Pivotal studies

<table>
<tr><td colspan="2">Title of study: 24-week double-blind, randomised, placebo-controlled, parallel group sequential study with sample size reassessment comparing the efficacy, safety and tolerability of two doses of teduglutide (0.05 mg/kg/day and 0.10 mg/kg/day) and placebo in subjects with Parenteral Nutrition-Dependent Short Bowel Syndrome</td></tr>
<tr><td colspan="2">Investigators (Study center): 32 centers, as previous pivotal</td></tr>
<tr><td>Studied period:
Recruitment may last from 33 to 44 months + 6 months follow-up of last randomised patient, if direct extrapolation from previous trials.</td><td>Phase of development:
Phase III</td></tr>
<tr><td colspan="2">Objectives
Primary: to evaluate the efficacy, safety, tolerability and pharmacokinetics of teduglutide compared with placebo in subjects with parenteral nutrition (PN)-dependent Short Bowel Syndrome (SBS).</td></tr>
<tr><td colspan="2">Design: Randomised, double-blind, placebo-controlled, parallel, group sequential and multicentre study aimed at evaluating the efficacy, safety, tolerability and pharmacokinetics of teduglutide in patients with PN-dependent SBS.
Run-in with optimisation of PN (3 days to 8 weeks) and stabilisation of PN (4-8 weeks); then stratified randomisation (by ≤6 or >6 L/week of PN at baseline) to parallel comparison of three groups (low dose, high dose and placebo) for 24 weeks; then continued in a separate extension study where patients in placebo will be crossed to active (randomization to low or high) and patients in active continued with same treatment for 28 additional weeks.
Randomisation will be balanced for treatment groups, site, and PN at two levels of consumption.
The study will be double blind, but swelling of stoma is as an AE to teduglutide that may break the blind. Persons responsible for assessment of PN requirements and adjusting doses of PN will be different from the ones conducting physical examinations and assessing safety.</td></tr>
</table>

Number of patients, by arm: 129 in total, 43 subjects in each of the teduglutide treatment groups and placebo group. **Intended sample size:** A maximum sample size of 43 patients per group (129 in total) would be needed to detect a difference of 220 mL/day between an active group and placebo in the change from baseline in PN volume, assuming SD of 475 in active and 200 in placebo. Since in average sequential trials allow for 30% sample size saving, there would be chances to finish with similar sample size as initially planned. Interim analysis are planned at 78 patients (60%) and 104 patients (80%), while reducing risk for a negative trial. Finally, a sample size reassessment is planned at the second interim analysis to cope with potentially deviation from variability in the study design. **Populations for analysis:** ITT including all subjects randomised
Main inclusion criteria: SBS due to the most common causes of SBS (surgical resections due to Crohn´s disease, cancer, vascular insufficiency and volvulus). Patients have to be dependent on PN at least 3 times weekly for at least 12 months. Prior to randomisation to study treatment patients have to be stable in their disease for the past 4 weeks as measured by usage/volume of PN, urinary output, urine sodium, renal function, haematocrit, and motility altering medications
Main Exclusion criteria: SBS patients with possible fluctuating activity in the disease (due to radiation therapy, active Crohn´s disease and celiac disease), history of cancer or clinically significant lymphoproliferative disease, history of alcohol or drug abuse (within previous year) and failure to washout of forbidden medications.
Test product, dose and mode of administration, duration: Teduglutide (0.05 or 0.10 mg/kg/day) was administered by the s.c. route once daily into one of the four quadrants of the abdomen or either thigh, for 24 weeks.
Reference therapy, dose and mode of administration, duration: **Active:** none **Placebo:** matching placebo for 24 weeks; teduglutide and placebo vials and their content were identical in appearance.
Criteria for evaluation (Definition, timing of assessments): **Efficacy assessments:** **Timing of assessments:** Weeks 0, 4, 8, 12, 16, 20, and 24 (main). **Main efficacy assessment:** Change from baseline in the "daily" PN volume for 14 days, calculated as [Sum (PN volume during the preceding 14 days)/(14 days)] at week 24. **Secondary variables:** Change from baseline in the "weekly" PN volume for 14 days, calculated as [Sum (PN volume during the preceding 14 days)/(14 days)] x 7 days at week 24. Relative change from baseline in the "weekly" PN volume for 14 days. N (%) of subjects who demonstrated a response (≥20% reduction from Baseline in weekly PN volume) at Week 20 maintained at Week 24; Ordered categorical (graded) criterion accounting for duration (weeks 16 and 20 and weeks 20 and 24) as well as intensity of response (percentage reduction in PN volume; 20-100%) meaning that subjects with larger, earlier and/or more sustained response had higher "weight" in the outcome (score of 0-5). A Goal Attainment Scale will be applied, by defining with each patient which is their goal for the treatment, and assigning 5 possible scores in which – 2 is for the least favorable outcome, 0 for the most likely treatment outcome, and +2 for the most favorable treatment outcome.

N(%) of subjects who achieved at least a 1-day reduction in weekly PN; the absolute reduction from baseline in weekly PN kilojoules; absolute reduction of weekly volume of PN from baseline; change from baseline in plasma citrulline at dosing week 24. Exploratory: Time to 20% reduction in PN volume, time to discontinuation of PN, time to a 1-day reduction in weekly PN, N(%) of subjects with reduced i.v. catheter access at week 24, change from baseline in bone markers BSAP and NTx, in lumbar spine ad hip BMD, and in PTH at week 24, change from baseline in QoL at weeks 4, 8, 12, 16, 20, and 24 (scores to be used were: SF-36, IBDQ, and abbreviated EuroQol)., the mucosal crypt-villus architecture and cellular composition within the small and large intestine. **Safety:** Laboratory safety samples were evaluated at least once during the seven days following a PN reduction, accompanied by determination of 48-hour urine output. In addition, evidence of dehydration, such as body weight, clinical signs and symptoms, was assessed day 3 to 4 and day 6-7 following a reduction. **Pharmacokinetic data / Pharmacodynamic data / other:**
Statistical methods **Analysis of efficacy:** Change from baseline in the "daily" PN volume for 14 days, calculated as [Sum (PN volume during the preceding 14 days)/(14 days)] at week 24 will be compared between groups using a Mixed Models for Repeated Measures (MMRM) test statistics adjusted for baseline values for PN volume and the randomisation stratification variables. Interim analysis will be done at the time when 60% (78) and 80% (104) of planned study patients will have reached their 24 weeks assessment. A sample size reassessment will be done in the second first interim analysis. The percent and absolute change in PN/i.v. volume from baseline to the last dosing visit will be compared between treatments using MMRM adjusted by baseline PN volume and accounting for stratification factors. Duration of response based on the number of consecutive visits (categorised as 0, 1, 2, and ≥3) for which the subject had a 20 to 100% reduction in weekly PN/i.v. volume from baseline at Week 24 plus previous scheduled visits with a 20 to 100% reduction from baseline, and N (%) of subjects with each level of the graded (or ordered categorical) response, will be compared using extended Cochrane – Mantel Haenzel test statistics (with standardized mid-ranks) adjusted for the randomisation stratification variable. **Missing data handling:** Sensitivity analyses using different ways of handling missing data will be applied to test if the results are or not sensitive to this factor. A worst scenario will be applied, where patients who discontinued the study prematurely will be assigned the score "0" for the response variable (i.e. no response), amongst others. **Safety:** descriptions will be done for AE, SAE and AE leading to discontinuation. A DSMB will be in place.

6.1.4 Supportive confirmatory efficacy data

Same as actual development

6.1.5 Supportive safety data

Same as actual development

6.2 Option 2: randomized withdrawal study

6.2.1 Safety and tolerability

Same as actual development

6.2.2 Proof of activity and exploratory dose-finding

Same as actual development

6.2.3 Pivotal study

A single pivotal trial with three phases: treatment phase with selection of responders, randomised withdrawal and a final phase of re-treatment of all patients not reaching response.

Title of study: A randomised, double-blind, placebo-controlled, randomised withdrawal parallel-group 24-Week study in patients with Parenteral Nutrition-Dependent Short Bowel Syndrome who respond to teduglutide therapy.	
Investigators (Study center): 27 sites	
Studied period: Recruitment may last 12 months + 6 months follow-up for last randomised patient, based on previous trials.	**Phase of development:** Phase III
Objectives **Primary:** To confirm the efficacy, safety, and tolerability of teduglutide compared with placebo in SBS subjects dependent on parenteral support (PN and/or i.v. fluids). **Secondary:** To describe the efficacy and safety of teduglutide in a number of secondary parameters related to PN and clinical signs and symptoms, and the impact on functionality and subject's qualify of life (QoL).	
Design: Double blind parallel groups randomised withdrawal study of three sequential phases: Prior to phase 1, patients will collect clinical information and PN requirements for 8 weeks to be used as their baseline assessment. - Phase 1: Open label treatment phase for selection of reponders for 24 weeks, with an standardised parenteral weaning algorithm aimed to reduce PN requirements at fixed schedule - Phase 2: patients who have already reached a reduction of the weekly requirements of PN by 20% from pre-study during the selection phase will be randomly assigned to one dose of tedugutide or placebo (1:1 ratio) during 24 weeks. Randomisation will be stratified by 2 levels of baseline PN/i.v. volume (≤6 L/week or >6 L/week). During this phase, patients increasing their requirements of PN by 50% during the randomised withdrawal phase will be rescued and re-treated with teduglutide; all others will complete 24 weeks of treatment. - Phase 3: an open label long-term treatment phase where patients will be treated with teduglutide for 2 years, in the scope of trial CL0600-021 (as in the original development).	
Number of patients, by arm: **Intended sample size:** In phase II a total of 15 patients per group would allow to detect significant differences in the percentage of patients not increasing their weekly PN volume requirements by 20% or more if these are about 70% in active and below 20% in placebo, or bigger differences, with α alpha=0.05, 2-sided test and power=90%, based on a Fisher's Exact	

Test. To achieve a sample size of 30 randomized patients that can be randomized into phase 2, and assuming a 60% of patients reaching the goal of 20% reduction (or >2L per week) of PN requirements during phase 1 (based on open label data), 50 patients should enter the Phase 1 to obtain 30 patients eligible for randomization. To that, and considering that 60% of screened patients are eligible to enter phase 1, based on previous studies (139 patients were screened to randomize 86 patients during 21 months), approximately 85 patients should be screened during approximately a period of 12 months. **Populations for analysis:** ITT will include all subjects randomised, and will be the primary study population, (with per protocol analysis supportive).
Main inclusion criteria: SBS due to the most common causes of SBS (surgical resections due to Crohn´s disease, cancer, vascular insufficiency and volvulus). Patients had to be dependent on PN at least 3 times weekly for at least 12 months. Prior to randomisation to study treatment patients had to be stable in their disease for the past 4 weeks as measured by usage/volume of PN, urinary output, urine sodium, renal function, haematocrit, and motility altering medications. For entering phase 2 of the study, patients will be required to have shown a 20% reduction of PN requirements from baseline after24 weeks open label treatment with teduglutide under a scheduled protocol for PN weaning.
Main Exclusion criteria: SBS patients with possible fluctuating activity in the disease (due to radiation therapy, active Crohn´s disease and celiac disease), history of cancer or clinically significant lymphoproliferative disease, history of alcohol or drug abuse (within previous year) and failure to washout of forbidden medications. For entering phase 2 of the study, patients will be excluded if they have not tolerated treatment, or have not responded to open label treatment with teduglutide after 24 weeks and have not achieved a 20% reduction of PN requirements, or >2L/ week, or bigger from baseline after24 weeks under an scheduled protocol for PN weaning.
Test product, dose and mode of administration, duration: Teduglutide (0.05 /kg/day) will be administered by the s.c. route once daily for 24 weeks in phase 1, as open—label treatment; for 24 additional weeks in the group randomized to active in phase 2, in blind conditions, and as an open-label treatment again in phase 3.
Reference therapy, dose and mode of administration, duration: **Active:** none **Placebo:** Placebo identical in appearance to teduglutide, administered with the same posology daily in double blind conditions during the phase 2. The study center personnel, the sponsor, and all personnel associated with the monitoring or data management for the clinical study will be blinded to the treatment assignment. However since the test drug may induce visible changes in the appearance of the intestinal mucosa, patients with jejunostomy/ileostomy may be aware of what kind of the treatment they receive.
Criteria for evaluation (Definition, timing of assessments): **Efficacy assessments:** **Timing of assessments:** Phase 1: Weeks 0, 4, 8, 12, 16, 20, and 24 (main)**.** Phase 2: Weeks 0, 4, 8, 12, 16, 20, and 24 (main)**. Phase 3:** monthly assessments. **Main efficacy assessment:** N (%) of subjects who do not increase their requirements for weekly PN volume by 20% or more, or >2L per week, at Week 24 of phase 2, as compared to day 0 of Phase 2.

Secondary variables: N (%) of patients requiring rescue during the double blind period because of increased PN requirements above 50% from day 0 of phase 2. Reductions in PN/i.v. volume or the direct effects of improved intestinal absorption of fluid; duration of response (ie., total number of weeks at ≥ 20% reduction from Baseline); the proportion of subjects with a ≥ 20% reduction or a ≥ 2 liter (L) reduction from Baseline in weekly PN at Week 20 and maintained through Week 24; N(%) of subjects who stopped PN and time of discontinuation; and absolute change and percent change in PN between Baseline and last dosing visit. Ordered categorical (or graded) response that accounts for both intensity and duration of the response at the end of the 24-week treatment period. Subjects' quality of life (QoL) using a subject-reported outcome SBS specific QoL scale. **Safety:** Adverse events, 12-lead ECG, vital signs, laboratory safety data, antibodies to teduglutide and/or E. coli protein (ECP), and changes in urine output and body weight.
Statistical methods **Analysis of efficacy:** Primary endpoint responder rates will be evaluated using Cochran–Mantel–Haenszel test statistics adjusted for the randomisation stratification variable (<=6 or >6 L/week of PN at baseline) for the proportion of patients not increasing their requirements during phase 2. Secondary variables will be analysed similarly if categorical, and through MMRM adjusted by baseline values at day 0 of phase 2 and stratification factors, if continuous variables. Plots for average values for weekly PN volume at week 0 and 24 in phase 1, 2 and 3 will be presented. Time to event will be analysed by log-rank tests and described by Kaplan –Meier. Similar analysis will be done in phase 3. Descriptions for the rate of responses and all parameters considered as secondary variables, including adverse events, will also be done for data collected during phase 1. **Missing data handling:** Sensitivity analyses using different ways of handling missing data will be done. Patients who discontinue the study prematurely will be assigned the score "0" for the response variable (i.e. no response). **Tolerability and acceptability:** descriptive data on AE, SAE and AE leading to discontinuation will be presented by phase and overall. A DSMB will be in place.

6.2.4 Supportive confirmatory efficacy data

Will correspond to Phase 1 and Phase 3 of the pivotal trial

6.2.5 Supportive safety data

Will be generated by open-label extensions of studies and compassionate use, if any.

7 Analysis of the practical, ethical and regulatory impact

7.1 Option 1: group sequential study with sample size reassessment and simultaneous stopping rule

Method assessed Option 1: group sequential study with sample size reassessment and simultaneous stopping rule	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements	No (similar)	The design may be similar or include more patients than the previous one, but the risk of study failure is smaller, so that the need to do the second trial may be avoided
• May shorten time to study completion	No (similar)	Will depend on final sample size, but the need for a second study will be avoided
• May ease recruitment	No (similar)	Will depend on final sample size, but the need for a second study will be avoided
Statistical assessment:		
• Improves internal validity	No (similar)	Previous study was already a double blind parallel trial
• Increases stability of estimates	No (similar)	Previous study was already a double blind parallel trial
• Increases sensibility to changes	Yes	Use of a continuous variable instead of binary increases sensitivity to change; simultaneous stopping rule allows to test all groups, with no hierarchical boundaries
• Compliant with predetermination	Yes	Predetermined interim analyses and stopping rules
• Consistency (discuss)	No	If completed earlier than anticipated, may reduce ability to test secondary parameters
• Robustness of method (discuss)	Yes	As described in article of Urach et al.
• Protection against type I and II errors (discuss)	Yes	As described in article of Urach et al.
Regulatory assessment:		
• Risk of bias and credibility	No (similar)	Previous study was already a double blind parallel trial
• External validity (discuss)	No (similar)	Same target population, potentially larger number of patients exposed
• Therapeutic positioning and comparisons	No	Compared to placebo, as main development
• Informative on relevance and clinical impact	Yes	Size of effect is descriptive allowing better interpretation, and GAS provides direct assessment of relevance by patients
• Enough information on safety	Yes	Larger number of patients treated within double-blind comparison than previous pivotal, although if second trial not required, the overall safety database is reduced
• Suitable information for risk-benefit balance	Yes	More precise information on treatment effect and information on clinical relevance than before

Method assessed Option 1: group sequential study with sample size reassessment and simultaneous stopping rule	Improves?	Comments
Ethical assessment:		
• May minimise risks	No (similar)	
• May maximise access to treatment	No (similar)	Similar setting as previous study; larger size of this pivotal trial, but second study will likely not be necessary
• May minimise unnecessary exposure to ineffective treatments or placebo	Yes	Sequential design may allow early interruption as soon as confirmation is achieved
• Considers patient input	Yes	Through GAS

The confirmatory program for Revestive was prolonged in time for almost 7 years due to the need to replicate the main pivotal trial, because the analysis for the main efficacy end-point was planned as a hierarchical sequence of testing, and since the high dose did not meet significance, the significant results observed for the low dose could not be deemed as confirmatory. The subsequent pivotal trial was positive for the low dose.

A multi-arm group sequential design is proposed as the first (and only, if positive) pivotal trial with a simultaneous stopping rule using either a continuous variable, which could be the change from baseline in PN volume, analysed by MMRM adjusted by baseline value, or the same variable used in the actual development ("weekly" PN volume for 14 days, calculated as [Sum (PN volume during the preceding 14 days)/(14 days)] x 7 days), but untransformed, instead of transformed to binary as in the previous trials (20% or more decrease).

These changes are deemed as acceptable from a regulatory point of view, and could have allowed the confirmation of efficacy, thus avoiding the need for a second trial. Interim analyses for futility/superiority are planned at 60% of sample size (78 patients) and after 80% of sample size (104 patients).

A sample size reassessment is proposed in the second interim analysis as a contingency in case that unexpectedly high variability will occur.

Finally, a GAS is introduced as secondary variable in order to support the assessment of clinical relevance of findings.

7.2 Option 2: randomised withdrawal study

Method assessed: Option 2: randomised withdrawal study	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements	Yes	Half the size for double-blind period
• May shorten time to study completion	Yes	Half the duration of the recruitment period
• May ease recruitment	Yes	More acceptable by patients since all will receive active
Statistical assessment:		

Method assessed: Option 2: randomised withdrawal study	Improves?	Comments
• Improves internal validity	Yes	Reduced heterogeneity because of enrichment method
• Increases stability of estimates	Depends	Smaller sample size, but smaller impact of heterogeneity due to enrichment
• Increases sensibility to changes	Yes	Enriched design allows to maximise treatment effect and thus increases sensitivity
• Compliant with predetermination	Yes	Conventional approach to hypothesis contrast
• Consistency (discuss)	No (similar)	May decrease sample size and thus ability to detect differences in secondary variables
• Robustness of method (discuss)	Yes	The sequence of events (treatment, withdrawal and retreatment) provides a strong evidence on causality
• Protection against type I and II errors (discuss)	Yes	Conventional statistics
Regulatory assessment:		
• Risk of bias and credibility	No	Enrichment design and selection of responders only may lead to overestimation of effect.
• External validity (discuss)	No	Subset of patients already responding, maybe only a small subset, not representative of the intended population.
• Therapeutic positioning and comparisons	Depends	Subset of patients already responding, may represents a small subset. However, is the subset of interest, so actually may help to position the treatment in clinics
• Informative on relevance and clinical impact	Yes	Useful information on response, on effects of withdrawal, and whether the reintroduction of treatment may allow recovering full effect of drug
• Enough information on safety	No	Smaller sample size
• Suitable information for risk-benefit balance	No	Lack of data on safety, and efficacy obtained may be potentially overestimated
Ethical assessment:		
• May minimise risks	Yes	Only patients obtaining benefit participate in full trial
• May maximise access to treatment	Yes	All patients are given the chance to check if they are responders
• May minimise unnecessary exposure to ineffective treatments or placebo	Yes	Placebo limited to a short period and rescues systems in place during the trial; a DSMB will be in place
• Considers patient input	No	No GAS is included because of the different phases

In summary, option 2 includes the use of an enrichment design with randomised withdrawal and rescue of failures, which may allow smaller sample size and shorter recruitment, as well as to likely observe bigger effect sizes.

Besides, the ethical assessment suggests that patients may perceive this design as less risky and more open to treatment access (always if assuming that the drug is having a beneficial effect, which may be often a bias). In any case, enhanced access to treatment represents incentives to recruitment; this may maximise the practical benefits of the design.

Such a design increases efficiency in study conduction but at the expense of smaller external validity and a potential overestimation of efficacy, which may add to regulatory uncertainties. Besides, the use of GAS is deemed inappropriate in this setting because most of the therapeutic improvement is anticipated to occur in an open label uncontrolled phase, so that no comparison may be available for interpretation of results. Yet, from a clinical perspective, the method provides an intuitive and clinically meaningful sequence of events and allows recommending a therapeutic approach to patient management.

8 Recommendations

The development of Revestive in the treatment of Short Bowel Syndrome is considered a representative model within the cluster of chronic stable conditions, since patients have a long lasting condition that has reached a reasonably stable status.

Heterogeneity across potential patients may be more important in this particular development than in other medical conditions within the cluster. This may also explain why the clinical program used strict inclusion and exclusion criteria. As a consequence, the duration of the trials was longer than expected for a prevalent condition, thus allowing consider methods using information as acquired – such methods may not be relevant in other medical conditions within the cluster.

The dose selection was based on a very small exploratory trial with only 16 patients, which was inconclusive due to heterogeneity and lack of standardised protocols for food intake. Actually, the first pivotal trial failed because of uncertainties in the dose selection strategy.

A multi-arm group sequential design is proposed as the first (and only, if positive) pivotal trial, testing two doses and placebo, with a simultaneous stopping rule using a continuous variable analysed by MMRM adjusted by baseline value. The method allows reducing the risk of failure due to uncertainties in dose selection based on few patients, and increases the sensitivity to changes because of using a continuous variable. May require higher sample size, but reduces substantially the risk of negative trial and thus the need for a second pivotal trial. Furthermore, a sample size reassessment is proposed in the second interim analysis as a contingency in case that unexpectedly high variability will occur.
An intermediate variable is available that is deemed reasonably representative of the clinical goal. However, care should be paid to the relaxation of the actual treatment goal when an intermediate variable is used. Changes that are significant in the intermediate variable are then discussed for relevance, and the initial treatment objective might be forgotten thereafter. Since the clinical relevance of the changes observed may be difficult to establish, and is

unlikely that hard outcomes, such as lower rates of complications of PN, or complete weaning of PN, a GAS introduced as secondary variable may add value to the data and reduce uncertainties in regulatory assessment.

A second option for development includes the use of an enrichment design with randomised withdrawal and rescue of failures, which may allow smaller sample size and shorter recruitment, as well as to likely observe bigger effect sizes. Such a design increases efficacy but at the expense of smaller external validity and a potential overestimation of efficacy. Yet, from a clinical perspective, the method provides an intuitive and clinically meaningful sequence of events and allows recommending a therapeutic approach to patient management.

In both scenarios, risks may include serious events, such as malignancy, that cannot be qualified with a clinical development perspective of 1 year (or less), unless their magnitude is outstanding. The risk/benefit assessment, thus, is uncertain in both the relevance of the clinical changes and the magnitude and severity of potential adverse reactions. The approach of continuous monitoring after commercialisation through registries, together with a risk minimisation strategy including alert in product information and thorough monitoring of patients in clinical practice, seems the best (only) option to manage this uncertainty.

Annex 4. 4. Soliris® (eculizumab)

Clinical Development Plan

Soliris® (eculizumab)

Clinical development for the long-term enzyme replacement therapy in patients with a confirmed diagnosis of aHUS disease

Version 1; 25/09/2017

1. Index

2. Introduction

2.1. Background

2.1.1. Disease and currently available alternatives

Paroxysmal nocturnal haemoglobinuria (PNH) is a rare blood disorder with high morbidity and mortality. PNH is clinically defined by the deficiency of the endogenous glycosyl phosphatidylinositol (GPI)-anchored complement inhibitory protein CD59 on the surface of blood cells. CD59 normally blocks the formation of the terminal complement complex (also called the membrane attack complex) on the erythrocyte surface, thereby preventing haemolysis. The pathophysiology of PNH is directly linked to the complement-mediated destruction of the susceptible PNH red blood cells, which results in intravascular haemolysis, the primary clinical manifestation in all PNH patients. PNH is a clonal acquired genetic disease arising from a somatic mutation in the gene pig-A, located in the X- chromosome. Inactivating mutations appear only in a proportion of cells (PNH cells) and this proportion can vary among patients and over time in a single patient.

The estimated prevalence of PNH is 13 cases per million. PNH is mostly a disease of adults, although childhood cases have been reported. The median age of onset is in the thirties. Patients have an approximately 15 year median survival from its initial diagnosis. PNH is associated with a wide range of clinical findings, several of which are potentially life threatening. The common clinical manifestations of PNH are haemolytic anaemia, venous thrombosis and deficient haematopoiesis. Many patients present with unexplained hemolytic anemia and associated symptoms, including fatigue, jaundice and red/pink urine. Excessive levels of cell-free plasma haemoglobin during intravascular haemolysis contribute to platelet activation, procoagulant activity and thromboembolism (TE), the leading cause of mortality in these patients (45%). Some patients have more unusual presentations in which nonspecific symptoms predominate, i.e. dysphagia, abdominal pain or erectile dysfunction due to nitric oxide depletion secondary to hemolysis. Intravascular hemolysis can lead to acute and chronic renal disease.

Hemolysis typically occurs at a low baseline level with periods of increased RBC lysis. Paroxysms of hemolysis at night may occur. Episodes of hemolysis may also be

triggered by various infectious or inflammatory stimuli including infections, surgery, strenuous physical activity, blood transfusions or alcohol use.

Anaemia is highly variable with haematocrit values ranging from ≤20% to normal. Red blood counts (RBC) are normochromic and normocytic unless iron deficiency has occurred from chronic iron loss in the urine. Granulocytopenia and thrombocytopenia are common and reflect deficient haematopoiesis. Clinical haemoglobinuria is intermittent in most patients and never occurs in some, but haemosidenuria is usually present.

There are no therapies specifically approved for the treatment of PNH and no generally applicable therapy adequately treats the serious conditions associated with PNH. The only curative treatment available to patients is bone marrow transplantation, which allows the replacement of the defective cells, however, this treatment is available for only a small proportion of patients since a suitable donor is required. Furthermore, transplantation may be associated with substantial risks. Current treatments for PNH are palliative and do not address the underlying disease process. Transfusion therapy is useful for raising the haemoglobin level and also for suppressing the marrow production of RBC during episodes of sustained haemoglobonuria. Iron replacement may be used, however it usually exacerbates haemolysis because of the formation of many new RBC, which may be sensitive to treatment. This may be minimized by giving prednisone (60 mg/d) or by suppressing the bone marrow with transfusions. Thrombolytic agents are used for acute thrombosis and antithymocite globulin is often used for treating the marrow hypoplasia.

2.1.2. Rationale for the development

The pathophysiology of PNH is directly linked to the complement-mediated destruction of the susceptible PNH red blood cells, which results in intravascular haemolysis, the primary clinical manifestation in all PNH patients.

Eculizumab is a humanized monoclonal antibody that binds to the human C5 complement protein. Eculizumab recombinant antibody inhibits C5 cleavage to C5a and C5b, preventing the generation of the terminal complement complex C5b-9 and thus blocking complement-mediated cell lysis and activation.

2.2. Scope of development

2.2.1. Target product profile

Indication	Soliris is indicated in adults and children for the treatment of Paroxismal nocturnal haemoglobinuria (PNH). Evidence of clinical benefit is demonstrated in patients with hemolysis with clinical symptom(s) indicative of high disease activity, regardless of transfusion history.
Route of administration	by intravenous infusion.
Pharmaceutical form	concentrate for solution for infusion
Posology	The recommended dose of Soliris for PNH is : Adults and children with body weight >40kg: 4-week initial phase (600mg every week- total 4 infusions) followed by maintenance phase (900 mg week 5th followed by 900 mg every 14+/-2 days) Children (<18 and <40 kg): body weight adjusted
Main target population	Patients with PNH with hemolysis with clinical symptom(s) indicative of high disease activity
Claims to be supported by the clinical development.	Treatment of haemolysis in PNH Reduction in transfusion requirements
Regions where the product should be marketed:	Global (or EU + USA)
Regulatory agencies that will be involved	EMA FDA

3. General investigational plan

3.1. Objective (s) of the development

The objective of the development is to provide pivotal support to the application for marketing authorization in the EU of the product, by generating:

- Confirmatory evidence of efficacy (superiority to placebo + SOC) in the control of hemolysis in patients with PNH disease.
- An appropriate safety database, including enough information to permit a characterization of the safety profile of the product, and a risk-benefit assessment of the product at the time of assessment of the marketing authorization application in the EU
- Any needed supporting information on the product to apply for a new drug

authorization in EU.

4. Assessment of applicability of methods

4.1. Representativeness of Soliris within the cluster

Soliris in the treatment of PNH has been chosen as a potentially representative example of Chronic Progressive led by one organ system within the Asterix clustering of medical conditions. Some aspects, however, should be considered regarding this example that may differ from other conditions in the cluster:

- The condition is characterized by a longer clinical course than acute conditions, and progresses during usually year(s), with an initial impairment of one system/organ, which may involve others along time. For this reason, start stop methods and intrasubject comparison are generally not feasible.
- Progressively reducing life quality and/or quantity of life: typically subjects are seriously disabled due to disease, but the degree of disability may vary from patient to patient depending on the size of the erythrocyte PNH clone. Parallel trials are needed when there are heterogeneity or poor predictability of clinical course. Designs must be add-on to SOC, due to severity. Enrichment designs may reduce heterogeneity.
- Due to the substantial potential impact on QoL, endpoints relying on patient reported outcomes, patient perceptions on the disease, disability and QoL may be relevant or required for decision-making.
- While in the cluster often disease assessment is dependent on patient inputs, with (time to change in) function(s) and QoL being key components of the efficacy measures, this is not the case for complement blockade in PNH, and in this sense the example is not representative. However, another characteristic within the cluster is the assessment of multiple end-points in the same domain, which is present in Soliris development.
- Since the condition is chronic with a relatively slow progression, it is acceptable to start research in adults where the disease is more prevalent, and then progress to the pediatric population once the proof of concept has been demonstrated.
- Because of the chronic course, recruitment will be rather based on prevalent cases than in incident cases. However, recrutiment may be difficult due to the low prevalence (13 in a million).
- There is a strong scientific rationale for the development of Soliris, due to the relevant role of the complement in the pathogenesis of the disease. Also, there is a good PD marker (haemolysis parameters) which may allow for intermediate analysis for the decision making.

4.2. Applicability of novel methodologies based on UMCU recommendations

- Applicable:
  - Long-short term outcomes: the effect on hemolysis may be a good biomarker, that can be validated along clinical development.
  - Fallback test for co-primary endpoints: already included 2 PEP, but uncertainties on the actual analysis conducted
  - Multi-arm group sequential designs with a simultaneous stopping rule UMW**. It requires a continuous endpoint.
  - Sequential design for small populations: Time needed recruitment vs. follow-up is >1. Quick response relative to recruitment. It requires a continuous endpoint
  - Bayesian sample size re-estimation using power priors. Bayesian approaches could be feasible in this case for implementing sample size reassessment or other adaptions, or for better integration of data increasing information and improving interpretation at the end of any study, specially if information on the condition is already existing and similar to the one obtained in the trial.
  - Dynamic borrowing through empirical power priors that control type I error
- Might be applicable:
  - Sample size reassessment and hypothesis testing in adaptive survival trials Phase II/III (seamless), might be an efficient approach, but "time to transfusion-event" analysis not necessarily translate automatically into a clinically relevant event (hemoglobin stabilization).
  - Simultaneous inference for multiple marginal GEE models
  - GAS could be applied, but not that relevant in this case, since objective end-points are feasible, and QoL and fatigue are already assessed and accounted for.
- Not applicable:
  - Meta-analytic approaches
  - delayed start randomization,
  - optimal exact tests for multiple binary endpoints,

Other considerations:

- At the time of the development, the only treatment options available were supportive /symptomatic plasma therapies, making RCT on top of BSC feasible.

- Some adaptions can be applied along the trial. Classical parallel sequential designs with long term comparison may be applicable, with early rescue / crossing over.
- Enrichment and stratification strategies were already applied. All the patients were randomly assigned in a 1:1 ratio to placebo or eculizumab by a centralized allocation method. Random assignment took place within 10 days of the qualifying transfusion in the observation period. Randomization was stratified according to the number of PRBC units transfused within 1 year prior to screening. The 3 randomization strata used were: between 4 and 14 units, inclusive, between 15 and 25 units, inclusive, and greater than 25 units.
- The recruitment time is almost equal to the time to outcome which would limit the potential advantage of sequential designs over a classically conducted clinical trial.

5. Actual development plan for Soliris in the treatment of PNH

Six clinical studies provide the basis for establishing the safety and efficacy of eculizumab therapy in the PNH patient population. These studies included 195 patients from 13 countries.

5.1. Safety and tolerability

No dedicated studies

5.2. Pharmacokinetics

No dedicated studies. PK and PD data obtained from the dose-finding, pivotal and extension studies.

5.3. Proof of activity/dose finding

Study C02-001, a phase 2 randomised open label pilot study of 12 weeks duration that enrolled 11 patients. This study had two study-specific extension studies (E02-001 and X03-001) totalling an additional 156 weeks. Data from the C02-001 open-label pilot PNH study showed that treatment with eculizumab at induction doses of 600 mg per week and maintenance doses of 900 mg every 2 weeks in PNH patients demonstrated statistically significant improvements in efficacy endpoints of serum lactate dehydrogenase (LDH) and units of PRBCs.

5.4. Pivotal study

- Study C04-001 was a randomised, double blind, placebo controlled phase 3 Study of 26-week duration, that enrolled a total of 87 patients

Title of study: C04-001 TRIUMPH Study: A Haemoglobin Stabilisation and Transfusion Reduction Efficacy and Safety Clinical Investigation, Randomised, Multi Centre, Double-Blind, Placebo-Controlled, Using Eculizumab in Paroxysmal Nocturnal Haemoglobinuria Patients

Investigators (Study center):

Studied period: 16 months First Patient Enrolled 27 Aug 2004 Last Patient Completed 27 Dec 2005	**Phase of development:** Phase III

Objectives
Primary: To evaluate the safety and efficacy of eculizumab in the study population, through both haemoglobin stabilisation and number of packed red blood cell (PRBC) units transfused.

Secondary: Secondary objectives were transfusion avoidance, haemolysis as measured by lactate dehydrogenase (LDH) area under the curve (AUC) during the treatment period from baseline to Visit 18, and Functional Assessment of Chronic Illness Therapy fatigue (FACIT-Fatigue) scale changes from Baseline to Visit 18. Exploratory endpoints included LDH changes from baseline to Visit 18, European Organisation for Research and Treatment of Cancer quality-of-life (QoL) questionnaire (EORTC QLQ-C30) changes from baseline to Visit 18, thrombosis, platelet activity, and measures of nitric oxide (NO) and free haemoglobin from baseline to Visit 18.

Design:
This was a phase III, randomised, multicentre, double blind, placebo-controlled clinical trial using eculizumab in PNH patients. All the patients were randomly assigned in a 1:1 ratio to placebo or eculizumab by a centralised allocation method. Random assignment took place within 10 days of the qualifying transfusion in the observation period. Randomisation was stratified according to the number of PRBC units transfused within 1 year prior to screening. The 3 randomisation strata used were: between 4 and 14 units, inclusive, between 15 and 25 units, inclusive, and greater than 25 units.

Number of patients, by arm: 88 randomised, 87 treated
Intended sample size: The sample size was selected to show that incidence rates of haemoglobin stabilisation during the treatment phase are 20% and 55% for the placebo and eculizumab groups, respectively. The median units of transfusion during the treatment phase were assumed to be 6 and 2 for the placebo and eculizumab groups, respectively. Based on these assumptions, with 35 patients per group and a 5% Type I error rate for each of the co primary endpoints, the study sample size of 75 patients was designed for approximately 82% power, using the 2-sided Fisher exact test for haemoglobin stabilisation and the Wilcoxon rank sum test for units of PRBCs transfused.
Populations for analysis For the co-primary endpoints and all secondary endpoints, the primary analysis was based on the ITT population.

Main inclusion criteria: age ≥18 years, patients must have required at least 4 episodes of transfusions in the 12 months prior to Visit 1 for anaemia or anaemia-related symptoms, patients must have a glycosylphosphatidylinositol-deficient red blood cell clone (type III cells) by flow cytometry of ≥10%, patients who were taking erythropoietin had to have been on a stable dose for 26 weeks prior to the screening visit (Visit 1) and the dose remained stable during the observation period and the treatment phase and patients must have had a platelet count of at least 100,000/mm3 either at Visit 1 or during the observation period.

Main Exclusion criteria: patients whose mean haemoglobin level prior to transfusion over the previous 12 months was greater than 10.5 g/dL and patients whose absolute neutrophil count of less than or equal to 500/µL.

Test product, dose and mode of administration, duration: Eculizumab patients received 600 mg of eculizumab IV once a week (within 5-9 days) for 4 doses, followed by 900 mg eculizumab IV 1 week later for 1 dose, then 900 mg eculizumab IV every 2 weeks (within 12-16 days). There was a total of 26 weeks of treatment.

Reference therapy, dose and mode of administration, duration: Placebo patients received placebo IV once a week (within 5-9 days) for 5 doses, then once every 2 weeks (within 12-16 days).

Criteria for evaluation (Definition, timing of assessments):
Efficacy assessments: The **co-primary endpoints** were haemoglobin stabilisation and units of PRBCs transfused during the treatment phase. For the haemoglobin stabilisation endpoint, patients that reached or dropped below their predetermined haemoglobin set point did not achieve haemoglobin stabilisation.
Secondary endpoints included transfusion avoidance, haemolysis as measured by the AUC of LDH, and QoL as measured by the FACIT-Fatigue scale. Exploratory endpoints included changes of LDH from Baseline to Visit 18, QoL changes as measured by the EORTC QLQ-C30 instrument, thrombosis rate, platelet activity, and free haemoglobin and NO measures.

Statistical methods
Analysis of efficacy: The primary analysis methods for the co-primary endpoints were the Fisher exact-test and the Wilcoxon rank sum test. The analysis of the secondary endpoint of transfusion avoidance was carried out using a 2-sided Fisher exact test. As a sensitivity analysis, those patients who dropped out of the study during the treatment phase prior to having a transfusion are classified as not requiring a transfusion. The AUC of LDH from Baseline to Visit 18 was presented for each patient and was analysed using the Wilcoxon rank sum test. The change of total FACIT-Fatigue scale score from Baseline was analyzed using a mixed-effects model with baseline as a covariate, treatment and time as fixed effects, and patient as a random effect. The exploratory endpoint of the changes of LDH from Baseline up to Visit 18 was analysed using a mixed-effects model, with treatment and time as a fixed effect and patient as a random effect. Also, LDH change from Baseline to Visit 18 was analysed using the Wilcoxon rank sum test.
Missing data handling: Not described
Tolerability and acceptability: Descriptive only

Main results (only actual scenario)
88 randomised, 87 treated, 85 completed 26-week treatment phase; 2 early discontinuations

Overview of Efficacy Endpoint Results

	C04-001		
	Placebo N = 44	**Soliris** N = 43	**P – Value**
Coprimary endpoints			
Percentage of patients with stabilised haemoglobin levels at end of study	0	49	< 0.001
PRBC transfused during treatment (median)	10	0	< 0.001
Secondary endpoints			
Transfusion Avoidance during treatment (%)	0	51	< 0.001
LDH levels at end of study (median, U/L)	2,167	239	< 0.001
LDH AUC at end of study (median, U/L x Day)	411,822	58,587	< 0.001
Free haemoglobin at end of study (median, mg/dL)	62	5	< 0.001
FACIT-Fatigue (effect size)		1.12	< 0.001

Haemoglobin Stabilization (ITT)

Randomization strata	Haemoglobin stabilization?	Eculizumab N = 43 n/N (%)	Placebo N = 44 n/N (%)	*P* Value[a]
Overall (N=87)	Yes	21/43 (48.8)	0/44 (0.0)	0.000000014
	No	22/43 (51.2)	44/44 (100)	
4 to 14 units (n=30)	Yes	12/15 (80.0)	0/15 (0.0)	0.000010521
	No	3/15 (20.0)	15/15 (100)	
15 to 25 units (n=35)	Yes	5/17 (29.4)	0/18 (0.0)	0.019061584
	No	12/17 (70.6)	18/18 (100)	
>25 units (n=22)	Yes	4/11 (36.4)	0/11 (0.0)	0.090225564
	No	7/11 (63.6)	11/11 (100)	

Note: Stabilization was calculated between Baseline and 26 weeks after first dose.
P values were calculated using Fisher's exact test.

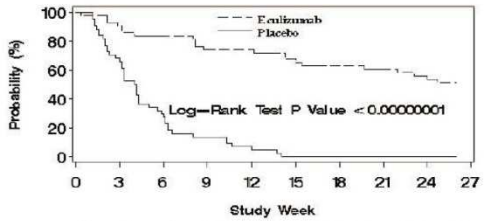
Secondary efficacy results
Transfusion avoidance was achieved in half of the patients treated with eculizumab while in none of those treated with placebo. This difference was observed independently of the stratum considered.

Summary of Units Transfused from Baseline to 26 Weeks (ITT)

Randomization strata	Eculizumab	Placebo	*P* value[a]
Overall (N)	43	44	<0.000000001
Mean (standard error)	3.0 (0.67)	11.0 (0.83)	
Median	0.0	10.0	
Range	(0.0, 16.0)	(2.0, 21.0)	
4 - 14 units (n)	15	15	0.000002311
Mean (standard error)	0.4 (0.29)	6.7 (0.72)	
Median	0.0	6.0	
Range	(0.0, 4.0)	(2.0, 12.0)	
15 - 25 units (n)	17	18	0.000665129
Mean (standard error)	4.2 (1.14)	10.8 (1.17)	
Median	2.0	10.0	
Range	(0.0, 15.0)	(2.0, 21.0)	
> 25 units (n)	11	11	0.000301977
Mean (standard error)	4.5 (1.59)	17.0 (1.04)	
Median	3.0	18.0	
Range	(0.0, 16.0)	(10.0, 20.0)	

P values were calculated using Wilcoxon's rank sum test.

Kaplan-Meier Plot of Time to First Transfusion During the Study



Treatment with eculizumab mitigates intravascular haemolysis, with median LDH AUC of 411,822 U/L x Day in placebo-treated patients and 58,587 U/L x Day in eculizumab-treated patients (P<0.001). P values <0.001 general and in all strata.
Assessments of functional status as the Fatigue Scale Scoring also shown relevant differences between groups.

AUC for LDH was also significantly lower in eculizumabgroup for all strata

Randomization strata	Eculizumab	Placebo	*P* value[a]
Overall (N)	43	44	<0.000000001
Mean (standard error)	3.0 (0.67)	11.0 (0.83)	
Median	0.0	10.0	
Range	(0.0, 16.0)	(2.0, 21.0)	
4 - 14 units (n)	15	15	0.000002311
Mean (standard error)	0.4 (0.29)	6.7 (0.72)	
Median	0.0	6.0	
Range	(0.0, 4.0)	(2.0, 12.0)	
15 - 25 units (n)	17	18	0.000665129
Mean (standard error)	4.2 (1.14)	10.8 (1.17)	
Median	2.0	10.0	
Range	(0.0, 15.0)	(2.0, 21.0)	
> 25 units (n)	11	11	0.000301977
Mean (standard error)	4.5 (1.59)	17.0 (1.04)	
Median	3.0	18.0	
Range	(0.0, 16.0)	(10.0, 20.0)	

P values were calculated using Wilcoxon's rank sum test.

Also for FACIT

Minimally Important Difference Change in FACIT-Fatigue Score at Week 26 (ITT)

Improved by at least 4 points?	Eculizumab n (%)	Placebo n (%)	*P* Value[a]
Yes	22 (53.66)	8 (20.51)	0.0028
No	19 (46.34)	31 (79.49)	

The *P* value was calculated using Fisher's exact test.

A reduction in intravascular haemolysis as measured by serum LDH levels was sustained for the treatment period and resulted in increased transfusion avoidance, a reduced need for RBC transfusion and less fatigue.

5.5. Supportive confirmatory efficacy and safety data

5.5.1. Single arm open label treatment

- Study C04-002, was a phase 3, open label study of 52-week duration that included 97 patients. The study was ongoing at the time of the MAA, with 26-week interim data available.

Title of study: C04-002 SHEPHERD Study: Safety in Haemolytic PNH Patients Treated with Eculizumab: A Multi-centre Open-label Research Design Study	
Investigators (Study center):	
Studied period: recruitment for 9 months, follow up for 6 months to end-point, total 15 months. First Patient Enrolled 15 Dec 2004 Last Patient Completed 26 weeks 21 Mar 2006 Last patient projected to complete treatment Sep 2006	**Phase of development:** Phase III
Objectives	

Primary:
The primary objective was to evaluate the safety of eculizumab in patients with at least one blood transfusion in the previous 24 months.
Secondary:

Design:
This was a phase III, open-label, multicentre study of eculizumab administered as an intravenous (IV) infusion to 97 PNH patients with haemolysis for a total of 52 weeks, with a pre-specified 26-week interim analysis. The study design consisted of a 2-week screening period and a 52-week treatment period with a prespecified 26-week interim analysis.

Number of patients, by arm:
Intended sample size: No formal sample size calculation. 97 patients enrolled
Populations for analysis

Main inclusion criteria: Patients ≥ 18 years who had received at least 1 blood transfusion in the previous 24 months were eligible. A PNH type III erythrocyte population ≥ 10%, platelets ≥ 30,000/mL, and lactate dehydrogenase (LDH) ≥ 1.5 times the upper limit of normal were also required as inclusion criteria.

Main Exclusion criteria: Patients who had received another investigational drug within 30 days of the first visit or had an absolute neutrophil count < 500/µL were excluded. Patients with complement deficiency, active bacterial infection, prior meningococcal disease, or prior bone marrow transplant were also excluded.

Test product, dose and mode of administration, duration:
During the induction period, patients received 600 mg of eculizumab administered by IV infusion once a week for 4 doses, followed by 900 mg eculizumab IV 1 week later for 1 dose. During the maintenance period, patients received 900 mg eculizumab IV every 2 weeks for a total of 52 weeks treatment. Each dose was administered by IV infusion over 25 to 45 minutes.

Reference therapy, dose and mode of administration, duration:
NA

Criteria for evaluation (Definition, timing of assessments):
Efficacy assessments: The **primary surrogate of efficacy** was haemolysis as measured by LDH area under the concentration curve (AUC).
The **secondary endpoints** were haemolysis as measured by LDH change from baseline and quality of life (QoL) as measured by the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue). Exploratory endpoints were Quality of life as measured by the European Organisation for Research and Treatment of Cancer Quality-of-Life Questionnaire Core 30 (EORTC QLQ-C30), thrombosis, platelet activity, nitric oxide (NO), and free haemoglobin measures.

Statistical methods Not detailed

Main results (only actual scenario)

Overview of Efficacy Results

	Soliris N = 97	P – Value
PRBC transfused during treatment (median)	0.0	< 0.001
Transfusion Avoidance during treatment (%)	51	< 0.001
LDH levels at end of study (median, U/L)	269	< 0.001
LDH AUC at end of study (median, U/L x Day)	-632,264	< 0.001
Free Haemoglobin at end of study (median, mg/dL)	5	< 0.001
FACIT-Fatigue (effect size)	1.14	< 0.001

* Results from study C04-002 refer to pre- versus post-treatment comparisons.

Treatment with eculizumab mitigates intravascular haemolysis, as measured by LDH AUC change from baseline (primary endpoint), with a median reduction in LDH AUC at 52 weeks of 632,264 U/L x Day (P<0.001).
Treatment with eculizumab mitigates intravascular haemolysis, as measured by median LDH change from baseline, with median LDH decreasing from 2,051 U/L pre-treatment to 269 U/L at 52 weeks (P<0.001). Assessments of functional status as the Fatigue Scale Scoring also showed relevant differences from baseline.

5.5.2. Open label extension of previous studies

Study E05-001: Phase 3b, open label 104-week extension of studies C04-001 (TRIUMPH), C04-002 (SHEPERD) and X03-001 (OLE of dose finding). This study included a total of 187 patients.

E05-001 is a phase III, open-label, extension study of eculizumab in patients with transfusion dependent, haemolytic PNH who have participated in the TRIUMPH (C04-001), SHEPHERD (C04- 002), or X03-001 Studies. Thus virtually all patients enrolled in eculizumab trials in PNH were entered in this extension protocol. TE events is the leading cause of morbid-mortality in patients with PNH.

The E05-001 statistical analysis plan prospectively identified that TE event rates would be analyzed as a cumulative event rate in the combined eculizumab-treated study population including data from all PNH trials (C02-001, E02-001, X03-001, C04-001, and C04-002). As was specified for all E05-001 planned comparisons, TE event rates during eculizumab treatment were compared to pre-eculizumab event rates in the same patients. However, it should be noted that findings from these studies were shown in non-controlled clinical trials.

Eculizumab treatment substantially reduced TE events in the combined eculizumab-treated PNH patient population. The reduction in TE event rate during eculizumab treatment as compared to individual matched patients pre-treatment was statistically significant (P<0.001).

Updated study E05-001 analysis of November 2006 data base lock of all eculizumab-treated patients which includes an additional 120 patient years of exposure shows that the reduction in thromboembolism risk has been maintained. When compared to the rate of thromboembolism events in all enrolled patients before treatment, eculizumab treatment results in a reduction in the thromboembolism event rate in the same patients in each of the individual clinical studies and a significant 7-fold reduction in the thromboembolism event rate overall from 7.37 events per 100 patient years pre-eculizumab treatment to 1.07 events per 100 patient years during eculizumab treatment (P<0.001).

E05-001; Comparison of Thrombosis/MAVE rates Across all Eculizumab PNH Studies: Total Pre-Eculizumab Treatment Rate vs. Eculizumab Treatment Rate

	C04-001	C04-002	C02-001/ E02-001/ X03-001	E05-001 All studies combined
Pre-treatment				
Patients (n)	43	97	11	195
Thrombosis events (n)	16	91	5	124
Patient years (n)	309.0	718.3	161.7	1683.4
Rate per 100 years	5.18	12.67	3.09	7.37
Eculizumab treatment				
Patients (n)	43	97	11	195
Thrombosis events (n)	0	2	0	3
Patient years (n)	21.8	96.9	34.2	281.0
Rate per 100 years	0.00	2.06	0.00	1.07[1]

[1] p = 9 x 10-14

5.6. Total patient exposure in the target indication

195 patients in clinical trials. Cumulative 264 patient-years during review of the dossier. Supportive safety data were obtained in 11 clinical studies that included 716 patients (492.20 patient years) exposed to eculizumab in six other indications.

5.8. Uncertainties/weaknesses identified

The demonstration of efficacy of Soliris in PNH patients with haemolysis was assessed in a randomized, double-blind, placebo-controlled 26 week study (C04-001). PNH patients were also treated with Soliris in a single arm 52 week study (C04-002) and in a long term extension study (E05-001).

The main limitations/uncertainties of the development program of Soliris in the treatment of PNH are related to the use of short-term surrogate endpoints, i.e. hemolysis, with some supportive clinical outcome data on fatigue, QoL. This was replicated in a second non-controlled clinical trial and the effect in the long-term where substantiated by the OLE study. Nevertheless, the assessment of efficacy in the prevention of TE events, which is the main cause of mortality in PNH, is based on non-controlled clinical trials.

No formal dose-finding studies were conducted. Dose selection was based on PK and PD data from the use of eculizumab in other non-related indications. Therefore, doubts remain on whether an alternative dose regimen might have been an optimal one.

There were also uncertainties related to the characterization of the safety profile of Soliris. Given the need to further collection of information of most important identified risks, and the need to complement global and long-term safety data, a safety registry was included as a FUM with clearly predefined timelines for the provision of the revised registry protocol and its implementation.

6. Alternative developments

6.1. Scenario 1: Multi-arm group sequential designs with a simultaneous stopping rule

This option was selected based on the results of the systematic exercise of applicability of methods (section 4.2), but also based on the clustering of conditions, reviewing the clinical characteristics of the intended therapeutic indication.

Option 1. Multi-arm group sequential designs with a simultaneous stopping rule.

6.1.2. Proof of activity

Equal to main development

6.1.3. Pivotal evidence

This method considers a multi-arm multi-stage design (3 arms or more, with provisioned interim analyses) for trials with a simultaneous stopping rule to achieve the objective to detect at least one efficacious treatment out of all tested arms and stops the conduct of the trial for all arms as soon as for at least one treatment arm efficacy is proven or not, when the critical boundaries of efficacy or inefficacy are crossed. When the trial is stopped early, the treatment efficacy testing is performed based on the interim data accumulated by then, and no additional patients are recruited. Using the simultaneous stopping rule, the critical boundaries are calculated more accurately and using this method, the probability of errantly concluding that one arm is efficacious when in fact it is not, is smaller. Therefore, this method can be efficiently applied if the primary aim of a trial in small populations with multiple treatment arms and with interim analyses is to prove that at least one of the arms is actually treating patients.

If two different treatment regimens would have been studied (for example in a dose-finding setting) and the two phase 3 studies merged, then it would have offered the possibility for MAMS with a simultaneous stopping rule. Taking into account that one of the main studies provisioned interim analysis in the design, then the application of a multi-arm multi-stage trial with a simultaneous stopping rule could have been possible to implement should there have been at least two treatment arms (e.g. two treatment regimens if the other study would have been merged);

One of the key criteria for applicability of this method is that the trial uses a continuous outcome and that it preserves cumulatively across all the interim analysis the maximum unwanted error to prove efficacy when in fact the treatment is not efficacious.

Furthermore, considering the above, this application would have plausibly offered room for more randomisation, stratification and minimisation strategies which could have been further applied to optimise the study.

Study outline

A single pivotal study, with multiple dose-regimens treatment arms is proposed as an alternative to the 2 phase 3 studies conducted.

Title of study: C04-001/2 TRIUMPH/SHEPERD Study: A Haemoglobin Stabilization and Transfusion Reduction Efficacy and Safety Clinical Investigation, Randomized, Multi Centre, Double-Blind, Placebo-Controlled, Using Eculizumab in Paroxysmal Nocturnal Haemoglobinuria Patients	
Investigators (Study center):	
Studied period: Expected 1.5 years, but may be reduced depending on the sequential results.	**Phase of development:** Phase III
Objectives **Primary:** The primary objective will the efficacy of eculizumab haemolysis as measured by lactate dehydrogenase (LDH) area under the curve (AUC) during the treatment period from baseline to Visit 18 **Secondary:** Secondary objectives were transfusion avoidance, Functional Assessment of Chronic Illness Therapy fatigue (FACIT-Fatigue) scale changes from Baseline to Visit 18. Exploratory endpoints included LDH changes from baseline to Visit 18, European Organisation for Research and Treatment of Cancer quality-of-life (QoL) questionnaire (EORTC QLQ-C30) changes from baseline to Visit 18, thrombosis, platelet activity, and measures of nitric oxide (NO) and free haemoglobin from baseline to Visit 18.	
Design: This was a phase III, randomized, multicentre, double blind, placebo-controlled clinical trial using eculizumab D1 and D2 in PNH patients using a multi-arm group sequential design with a simultaneous stopping rule. All the patients will be randomly assigned in a 1:1:1 ratio to placebo or eculizumab D1 or D2 by a centralized allocation method. Random assignment took place within 10 days of the qualifying transfusion in the observation period. Randomization will be stratified according to the number of PRBC units transfused within 1 year prior to screening. The 3 randomization strata will be: between 4 and 14 units, inclusive, between 15 and 25 units, inclusive, and greater than 25 units. Two Interim analysis for overwhelming efficacy are planed at the 60% and 80% of the overall sample size.	
Number of patients, by arm: 44 patients per arm (44:44:44) **Intended sample size:** The actual development plan was overpowered for the efficacy assessments, and post-hoc sample size calculations for any of the primary or secondary endpoint would lead to a much lower sample size (20 patients per arm or less). The current design will assume the uncertainties for the original design and will be based on the same sample size, the impact of this fact is discussed in sections 7-8.	

Populations for analysis The primary analysis was based on the ITT population.
Main inclusion criteria: The main inclusion criteria were age ≥18 years, patients must have required at least 4 episodes of transfusions in the 12 months prior to Visit 1 for anaemia or anaemia-related symptoms, patients must have a glycosylphosphatidylinositol-deficient red blood cell clone (type III cells) by flow cytometry of ≥10%, patients who were taking erythropoietin had to have been on a stable dose for 26 weeks prior to the screening visit (Visit 1) and the dose remained stable during the observation period and the treatment phase and patients must have had a platelet count of at least 100,000/mm3 either at Visit 1 or during the observation period. The main inclusion criteria were patients whose mean haemoglobin level prior to transfusion over the previous 12 months was greater than 10.5 g/dL and patients whose absolute neutrophil count of less than or equal to 500/µL.
Main Exclusion criteria:
Test product, dose and mode of administration, duration: The treatment will be placebo or eculizumab D1 and D2, in a double-blinded manner. D1 Eculizumab patients received 600 mg of eculizumab IV once a week (within 5-9 days) for 4 doses, followed by 900 mg eculizumab IV 1 week later for 1 dose, then 900 mg eculizumab IV every 2 weeks (within 12-16 days). D2 Eculizumab patients received 900 mg of eculizumab IV once a week for the first 4 weeks, followed by 1200 mg for the fifth week, followed by 1200mg every 2 weeks. There was a total of 26 weeks of treatment.
Reference therapy, dose and mode of administration, duration: Placebo patients received placebo IV once a week (within 5-9 days) for 5 doses, then once every 2 weeks (within 12-16 days).
Criteria for evaluation (Definition, timing of assessments): **Efficacy assessments:** Main outcome: Hemolysis measured LDH area under the concentration curve (AUC) **Secondary endpoints** Key secondary end points: (i) haemoglobin stabilization and (ii) units of PRBCs transfused during the treatment phase. Other secondary end points: transfusion avoidance, haemolysis as measured by changes in LDH from baseline, and QoL as measured by the FACIT-Fatigue scale. Exploratory endpoints included changes of LDH from Baseline to Visit 18, QoL changes as measured by the EORTC QLQ-C30 instrument, thrombosis rate, platelet activity, and free haemoglobin and NO measures.
Statistical methods **Analysis of efficacy:** The primary analysis will be based on the ITT population using an ANCOVA analysis. Two Interim analyses are planned at 60% and 80% of the overall sample size using

O'Brian-Fleming type boundaries for overwhelming superiority at a two-sided 5% type I error, using multi-arm group sequential design with a simultaneous stopping rule approach.

The analysis of the secondary endpoint of transfusion avoidance will be carried out using the 2-sided Fisher exact test and the Wilcoxon rank sum test. As a sensitivity analysis, those patients who dropped out of the study during the treatment phase prior to having a transfusion are classified as not requiring a transfusion. The FACIT-Fatigue scale collected secondary endpoint QoL data and will be scored according to the scoring guideline for this instrument. The change of total FACIT-Fatigue scale score from Baseline will be analysed using a mixed-effects model with Baseline as a covariate, treatment and time as fixed effects, and patient as a random effect. The changes of LDH from Baseline up to Visit 18 will be analysed using a mixed-effects model, with treatment and time as a fixed effect and patient as a random effect. Also, LDH change from Baseline to Visit 18 was analysed using the Wilcoxon rank sum test.

Missing data handling:

Worst Observation Carry Forward (WOCF) will be proposed as main missing data handling for primary outcome. Sensitivity analyses will comprise mixed models and Last Observation Carry forward (LOCF).

Tolerability and acceptability: Descriptive only

6.1.4. Supportive confirmatory efficacy/safety data

Same as previous: OLE studies

6.2. Scenario 2

Based on Asterix novel developed methods, a prospectively defined **meta-analysis of small trials** will be planned. This is not meant to be the basis for the MAA but to provide key data regarding the incidence of thrombo-embolic events (TEE), which are main cause of mortality associated to PNH. Without this information on TEE it is not possible to assess the long-term impact of the treatment on a hard and a fully meaningful end-point. This scenario might be a good complementary approach to scenario 1

Option 2: Prior distributions for variance parameters in sparse-event meta-analysis

6.2.1.First in man administration

Same as main development

6.2.2. Proof of activity and dose finding

Equal to main development

6.2.3. Pivotal evidence

Same as main development.

6.2.4. Supportive confirmatory efficacy/safety data

A prospectively defined meta-analysis of small trials will be planned to generate more robust information on the actual treatment effect on the reduction of the thrombo-embolic events (TEE) incidence. TEEs are critical and main cause of mortality associated to PNH. Also, meta-analytic techniques may be applied to the analysis of safety information, considering not only the PNH, but also other indications.

Data will be used from the pivotal trial and the open label trial, as well as from the open-label extension studies. Different approaches for handling small trials regarding the points of use of priors, sparse-event data, and heterogeneity estimators in zero cells meta-analysis, might be used. We propose to use two techniques developed within the Asterix project, for the primary analysis the "*Prior distributions for variance parameters in sparse-event meta-analysis*" and "*Heterogeneity estimators in zero cells meta-analysis*" as a complementary tool.

This approach will allow the fast development based on optimized approaches (like the one proposed in section 6.1, with the added value of assessing a harder end-point, to be conducted during the development and likely finalized early after the registration using information from post-authorization studies. This might eventually be complemented with sequential techniques to further control the type I error and thus be assessed on a sequential basis more than after a fixed amount of information.

Justification for the application of the primary analysis of the method "*Prior distributions for variance parameters in sparse-event meta-analysis*" approach. The small sample sizes in rare diseases make it particularly valuable to pool the data of small studies. When the primary outcome is binary, small sample sizes increase the chance of observing zero events. The frequentist random-effects model is known to induce bias and to result in improper interval estimation of the overall treatment effect in a meta-analysis with zero events. Bayesian hierarchical modelling could be considered as a promising alternative. Bayesian models are known for being sensitive to the choice of between-study variance (heterogeneity) prior distributions in sparse settings. In a rare disease setting, only limited data will be available to base our prior on.

Since the estimated prevalence of PNH is only 13 cases per million it will not be feasible that the meta-analysis will provide sound confirmatory evidence for the prevention of TEE, but at least a rough valuable estimate will be available. In fact, this study will provide useful complementary information to efficacy variables (such those

used in the actual pivotal studies: rates of haemoglobin stabilization, the median units of transfusion, DH area under the concentration curve (AUC)), and in particular with regards to the assessment of benefit risk taking into account the overall picture of efficacy and safety assessed in the meta-analysis, and potential relevant subgroups.

7. Analysis of the practical, ethical and regulatory impact

7.2 multi-arm multi-stage trial with a simultaneous stopping rule

Method assessed: Option 1: multi-arm multi-stage trial with a simultaneous stopping rule	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements	Yes	By merging the two phase 3 studies. Also, the sample size was overpowered in the original plan was. Thus, given that we have maintained the initial uncertainty (i.e. same sample size per arm), this approach will very likely lead to early termination.
• May shorten time to study completion	Yes	Yes, theoretically
• May ease recruitment	Yes	Options to receive active are bigger, so may stimulate participation
Statistical assessment:		
• Improves internal validity	Similar or even Better	Better if dose response is met
• Increases stability of estimates	No	Probably not, in fact the decision based on the interim analysis may be subject to less stability in the estimates
• Increases sensibility to changes	No	Contrast similar to conventional parallel
• Compliant with predetermination	Yes	Yes, interim pre-planned and methods for adjustment in place
• Consistency (discuss)	No	The ability to show an effect in secondary endpoints may be reduced if a reduced sample size or shorter term FU

Method assessed: Option 1: multi-arm multi-stage trial with a simultaneous stopping rule	Improves?	Comments
• Robustness of method (discuss)	Yes	Based on pre-established criteria for dose-selection, but may hamper observing an effect in relevant secondary clinical outcomes
• Protection against type I and II errors (discuss)	Similar	Methods for adjustment in place
Regulatory assessment:		
• Risk of bias and credibility	No	Parallel double blind design, appropriate methods for analysis
• External validity (discuss)	No	In fact, by merging the two studies into one with an enriched population, external validity might be hampered. A potential sample size reduction due to early termination might be another element in this regard.
• Therapeutic positioning and comparisons	No	Compared only with placebo
• Informative on relevance and clinical impact	No	In fact, PEP will be limited to a PD marker of hemolysis with no clinical outcome data as part of the PEP
• Enough information on safety	Yes	Little impact on the overall development, and long-term exposure will remain unmodified.
• Suitable information for risk-benefit balance	Better	Better information on dose selection, dose-response as an additional evidence of efficacy
Ethical assessment:		
• May minimise risks	Yes	By reducing the number of patients exposed to both placebo/active
• May maximize access to treatment	No	Randomization rate to active:placebo 2:1, but half of active patients assigned to a dose different form the selected one.
• May minimise unnecessary exposure to ineffective treatments or placebo	Yes	Yes, only 1/3 of the patients will receive placebo. This might be improved by reducing sample size in an early termination
• Considers patient input	No	None

By merging the two phase 3 studies and applying the multi-arm multi-stage trial with simultaneous stopping rules it can be anticipated some practical and ethical advantages, i.e. reducing sample size, time to completion and facilitating enrolment. The quality of the evidence is not affected in a relevant way and from the regulatory point of view the main advantage would be that this Option 1 may address one of the uncertainties of the actually conducted development plan by testing two different dose levels. However, it may negatively affect the ability of to provide evidence in clinical outcomes by reducing sample size, shortening duration and relying purely on PD markers. Therefore, this option is not considered to be of true added value, given that on major safety concerns existed on the dose tested and finally recommended for use.

7.3 Complementary design. Meta-analysis of small trials, prior distributions for variance parameters in sparse-event meta-analysis

Method assessed: Option 2: Complementary design. Meta-analysis of small trials, prior distributions for variance parameters in sparse-event meta-analysis	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements	Yes	Not for the MAA but yes for the total required evidence by incorporating data from several studies
• May shorten time to study completion	No	No
• May ease recruitment	No	NA
Statistical assessment:		
• Improves internal validity	Similar	Actually NA, but heterogeneity may be assessed
• Increases stability of estimates	Yes	More data by incorporating data from several studies
• Increases sensibility to changes	Yes	More data by incorporating data from several studies
• Compliant with predetermination	Yes	Planned Upfront
• Consistency (discuss)	Yes	By increasing the chances to show an effect on clinical outcomes
• Robustness of method (discuss)	Yes	Based on selection of clinical outcomes not just PD or surrogates
• Protection against type I and II errors	Similar	Methods for adjustment in place

Method assessed: Option 2: Complementary design. Meta-analysis of small trials, prior distributions for variance parameters in sparse-event meta-analysis	Improves?	Comments
(discuss)		
Regulatory assessment:		
• Risk of bias and credibility	No	Depending on the individual study designs
• External validity (discuss)	Yes	By incorporating different studies from different sources and handling heterogeneity
• Therapeutic positioning and comparisons	No	Compared only with placebo,
• Informative on relevance and clinical impact	Yes	Improved as PEP based on a clinical outcome
• Enough information on safety	Yes	Little impact on the overall development
• Suitable information for risk-benefit balance	Yes	Better information on clinical outcomes
Ethical assessment:		
• May minimise risks	No	No impact
• May maximize access to treatment	No	No impact
• May minimise unnecessary exposure to ineffective treatments or placebo	No	No impact
• Considers patient input	No	No impact

In summary, a prospectively defined meta-analysis of small trials will generate more robust information on the actual effect on of treatment on reducing the incidence of thrombo-embolic events (TEE) which are the critical and main cause of mortality associated to PNH. Also, meta-analytic techniques may be applied to the analysis of safety information, considering not only the PNH, but also other indications. Bayesian approaches could be a suitable option for better integration of data increasing information and improving interpretation at the end of any study, specially if information on the condition is already existing and similar to the one obtained in the trial. This would allow focusing on relevant outcomes, i.e. TEE, and overall efficacy/safety assessment, in addition to PD markers, which may increase robustness and relevance of the

development, by addressing one of the identified drawbacks of the actual development. No major ethical, practical impact, no sample size reduction at the time of the MAA, but much higher level of evidence for hard endpoints. Overall this approach is considered to add value to the development program.

8. Recommendations

The development of Soliris in the treatment of PNH is considered a representative model within the cluster of chronic progressive conditions led by one organ/system. There is a good scientific rationale for the development of Soliris, a complement inhibitor, in the treatment of PNH given the known pivotal role of the complement. Therefore, previous considerations on applicability of novel methods can reasonably be considered suitable options for conditions belonging to this same cluster.

In particular, parallel trials are needed due to the heterogeneity or poor predictability of clinical course, usually with add-on to SOC. Enrichment designs may reduce heterogeneity.

Given the relatively low progression of the disease and the lack of optimal curative treatments, there is a high willingness to accept trials even if SOC available. Unbalanced randomization to minimize risks may be useful to facilitate recruitment.

Disease assessment often highly dependent on patient inputs, with (time to change in) function(s) and QoL being key components of the efficacy measures; multiple end-points usually in the same domain may be acceptable/required. Often using surrogates that allow early (interim) results used for decision making.

When severe, classical parallel sequential designs with long term comparison may not be applicable, unless early rescue / crossing over. Therefore, some adaptions can be applied along the trial. The recruitment time is almost equal to the time to outcome which would limit the potential advantage of sequential designs over a classically conducted clinical trial. Nevertheless, multi-arm group sequential designs with a simultaneous stopping rule UMW might be a suitable option with practical advantages, i.e. reduces sample size and time to completion.

In particular, methods using prior information for meta-analysis are highly valuable options for small clinical trials with sparse events. Integration of data will increase key information on hard clinical endpoints as well as an integrated assessment of benefit/risk and identification of key subgroups. This may increase the relevance/robustness to the evidence generated. Of note, this design, when upfront planned, might be a complementary strategy for many other situations.

Annex 4. 5. Fabrazyme® (agalsidase beta)

Clinical Development Plan

Fabrazyme® (agalsidase beta)

Clinical development for the long-term enzyme replacement therapy in patients with a confirmed diagnosis of Fabry disease

Version 1; 23/09/2017

Index

2. Introduction

2.1. Background

2.1.1. Disease and currently available alternatives

Fabry Disease is a rare X-linked recessive glycosphingolipid storage disorder that is caused by deficient activity – subnormal or absent - of the lysosomal enzyme, α-galactosidase A. It is an extremely rare disorder, with an estimated prevalence of 500 to 1000 patients within the EU.

The absence of the enzyme leads to progressive accumulation of neutral glycosphingolipids, predominantly ceramide trihexoside (CTH), in most tissues and cell types, particularly the vascular endothelial and smooth-muscle cells leading to ischaemia and infarction. Fabry Disease is a heterogeneous multisystemic disorder with variable onset of symptoms affecting the nervous system, kidneys, heart, skin and gastrointestinal system. There are also atypical variants for which either renal or cardiac involvement is the only problem. It is not possible to predict the phenotype based on current knowledge of the different genotypes described for this disease.

The presentation of the disease is highly heterogeneous. Early manifestations in childhood or adolescence may include crises of excruciating pain in the extremities, angiokeratoma, hypohydrosis and corneal and lenticular opacities. In the second or third decade of life, early renal accumulation of glycosphingolipids has been observed throughout the renal vessels and glomeruli, resulting on the onset of renal insufficiency. Premature death usually occurs in the fourth or fifth decade of life and results from renal, cardiac, or cerebrovascular complications. Approximately one-third of heterozygous female carriers of Fabry Disease ultimately become symptomatic.

At present there is no specific curative treatment for the condition and patient management is limited to symptom control and supportive measures.

2.1.2. Rationale for the development

Fabrazyme is a recombinant human α-galactosidase (r-hαGAL), INN: agalsidase beta, which is produced by genetically engineered Chinese Hamster Ovary (CHO) cells. Recombinant hαGAL is a highly purified recombinant form of the naturally occurring human lysosomal hydrolase enzyme responsible for the metabolism of globotriaosylceramide (ceramide trihexoside; CTH; GL-3).

The rationale for enzyme replacement therapy (ERT) is to restore a level of enzymatic activity sufficient to hydrolyse the accumulated substrate. After administration, agalsidase beta is rapidly removed from the circulation and taken up by vascular endothelial and parenchymal cells into lysosomes, likely through the mannose-6-phosphate, mannose and asialoglycoprotein receptors.

2.2. Scope of development

The product development plan is limited to the studies providing clinical support to Fabrazyme as a long-term enzyme replacement therapy in patients with a confirmed diagnosis of Fabry disease.

2.2.1. Target product profile

Indication	Long-term enzyme replacement therapy in patients with a confirmed diagnosis of Fabry disease
Route of administration	Intravenous infusion
Pharmaceutical form	Powder for concentrate for solution for intravenous infusion (should be reconstituted and diluted prior to infusion)
Posology	1 mg/kg body weight, intravenously once every 2 weeks
Main target population	Patients with inherited deficiency in the activity of the lysosomal enzyme α-galactosidase A and Fabry disease
Claims to be supported by the clinical development.	Reduction of substrate (GL-3) accumulation, improvement of clinical signs and symptoms.
Regions where the product should be marketed:	Global (or EU + USA)
Regulatory agencies that will be involved	EMA FDA

3. General investigational plan

3.1. Objective (s) of the development

The objective of the development is to provide pivotal clinical support to the application for marketing authorisation in the EU of the product, by generating:

- Confirmatory evidence of dose-response and/or superiority to placebo in the long-term ERT for patients with Fabry disease.
- An appropriate safety database, including enough information to permit a characterisation of the safety profile of the product.
- Any needed supporting information on the product to allow a proper risk-benefit assessment of the product at the time of assessment of the new drug marketing authorisation application in the EU.

4. Assessment of applicability of methods

4.1. Representativeness of Fabrazyme within the cluster

Fabrazyme has been chosen as a potentially representative example of Chronic Progressive multidimensional condition within the Asterix clustering of medical conditions. This is a chronic condition with a relatively slow progression, representative of many other conditions within the cluster. Some aspects, however, should be considered regarding the general applicability of conclusion within this example that may differ from other conditions in the cluster:

- Because of slow progression, there is no emergency in access, and thus, research can start in adults to obtain proof of concept before studies are initiated in the pediatric population applying criteria of prudency. This may facilitate conduct of clinical trials. However, if the condition is extremely rapid in progression or treatment delay has a substantial impact on prognosis, paediatric trials or trials including paediatric patients may be justified as the first approach.
- Recruitment will generally be based on prevalent cases, thus already available cases, so theoretically can be done quickly and then adaptions and sequential methods may be less sensible. However, when the condition has very low prevalence and there is dispersion of patients (ultra-rare condition, 500-1000 patients in EU), the uptake of sites may be progressive, and then such methods may be applicable.

At the time of the development of Fabrazyme there were not effective SOC treatments available, so that no active control was identified. Patient management was limited to symptom control and supportive measures.

There is a strong scientific rationale for the development of Fabrazyme since the pathophysiology of the disease is well known: heterogenous multi organ/system damage due to substrate accumulation (sphingolipids, the most important GL3) secondary to an enzyme deficiency (alfa-galactosidase). A product for enzyme replacement therapy (ERT) is already authorised and available; Fabrazyme is also an ERT.

There is a direct pharmacodynamic marker related with the physiopathology of the disease, which may allow for intermediate analysis for the decision making: GL3 (sphingolipids) accumulation in different organs is the cause for the disease and its clinical presentation. Clearance of the substrate from different tissue organs, although not fully conclusive of efficacy, indicates that a clinical improvement or stabilisation among patients is to be expected.

Although feasible, no prospective registries were available at the time of the MAA, which might have been an important tool to consider in the plan.

4.2. Applicability of novel methodologies based on UMCU report

Applicable methodologies:

Long-short term outcomes: accumulation of the substrate in plasma and tissues is a good short-term surrogate marker, given that this is the underlying cause of the organ/system damage. This marker may be validated along clinical development. Already applied in Fabrazyme development (phase II dose-finding: plasma changes in GL3)

Sequential design for small populations: Time needed, recruitment vs. follow-up, disease features such as progression, return to baseline after treatment. Quick

response relative to recruitment. Sequential designs for small populations if the principal end-point (PEP) is replaced by a continuous endpoint might also be an applicable approach, but likely may not represent any added value.

Bayesian sample size re estimation using powers prior. Previous information on the clinical course can be suitable for Bayesian approaches and planning of adaptions. Bayesian approaches could be feasible in this case for implementing sample size reassessment or other adaptions, or for better integration of data increasing information and improving interpretation at the end of any study, especially if information on the condition is already existing and similar to the one obtained in the trial.

Candidate for GAS: Heterogeneous disease course/heterogeneous population or unstable baseline, measurement at functional level relevant.

Minimisation or stratification strategies would have been allowed based on the condition and the design of the study.

Might be applicable:

Multi-arm group sequential designs with a simultaneous stopping rule: Taking into account that the dose-response study previously conducted was deemed as acceptable but conditioned by further investigations, and given the potential applicability of sequential designs as mentioned above, then the application of a multi-arm multi-stage trial with a simultaneous stopping rule could have been possible to implement.

There should have been at least two treatment arms (e.g. two treatment regimens) but we conclude that the method is not immediately applicable. Considering the above, this application would have plausibly offered room for randomisation and/or minimisation, and stratification strategies which could have been further applied to optimise the study.

Dynamic borrowing through empirical power priors that control type I error: To be applied, good quality historical data should have been available. If two pivotal studies can be done, then dynamic borrowing could be applied to the analysis of the second trial, but gain is not anticipated to be relevant.

Not applicable:

Delayed start randomisation: due to expected long term effects and lack of relevance of early changes.

Sample size reassessment and hypothesis testing in adaptive survival trials: because variables are often not suitable to survival designs and adaptions are less efficient in chronic conditions recruiting prevalent cases.

Fallback tests for co-primary endpoints: in this case, this is concluded as not applicable from statistical point of view because a single primary was used in the actual development.

Optimal exact tests for multiple binary endpoints: in this case, this is concluded as not applicable from statistical point of view because a single primary was used in the actual development.

Additional considerations based on disease clustering:

- **Multiple end-points:** The use of methods based on multiple endpoints is considered as not applicable based on the endpoints actually chosen in the Fabrazyme development plan. However, given the involvement of multiple

organs in Fabry´s disease and the heterogeneity of the condition, methods based on relevant multiple end-points that may reduce the risk for failure due to choice of an insensitive primary variable, like the Fallback tests for co-primary endpoints and the optimal exact tests for multiple binary endpoints, might be applicable and may add substantial value (more complete assessment of the multidimensional nature of the disease) to the development plan in Fabry´s disease.

- **Enrichment:** Any method may be coupled with enrichment of study population through inclusion and exclusion criteria, as a mean to maximise the sensitivity of the sample to detect treatment effects. Thus, patients with clinical symptoms and/or functional impairment can be included allowing to combine histological changes with clinical outcome measures (functional, QoL, symptomatic changes), providing a more convincing (clinically relevant) demonstration of efficacy.
- **Registries and patient's associations:** Prevalent cases population may be identified from registries and patient's associations, speeding recruitment, and thus making sequential approaches and adaptions less efficient.
- **Parallel designs and stratification:** Parallel designs will be generally needed, due to progression and intersubject variability; for the same reason, stratification and prospective anticipation of subgroups are required to handle heterogeneity. Enrichment strategies commented above may also be useful to control heterogeneity.
- **Survival trials are not a suitable option** mainly because of the usually long-lasting course of the disease, but also because of the heterogeneity of the disease and difficulties to choose an acceptable primary end-point suitable for definition of survival parameters.

5. Actual development plan for Fabrazyme

The development was focused on the ability to reduce the accumulation of substrate through the long-term enzyme replacement therapy in patients with a confirmed diagnosis of Fabry disease (α-galactosidase A deficiency). A summary from the EPAR is included below.

5.1. Safety and tolerability

No dedicated studies.

5.2. Pharmacokinetics

No dedicated studies. Pharmacokinetic (PK) and pharmacodynamic (PD) data were obtained from the dose-finding, pivotal and extension studies.

5.3. Proof of activity/dose finding

Study FB9702-01 (US), a phase I/II supportive, dose- finding, single-centre, safety study, conducted in 15 patients testing 5 groups of doses: every other week: 0.3, 1.0, 3.0, or every other day: 1.0, 3.0. A total of 5 infusions. PK/PD data also were obtained.

5.4. Pivotal evidence

AGAL-1- 002-98 (US, Europe), a phase III Pivotal randomised, double blind, placebo controlled multi-centre study, conducted in 58 patients, randomized to receive 0 or 1q2w, for up to 20 wks.

5.5. Supportive confirmatory efficacy and safety data

AGAL- 005-99, a phase III Open label extension study, with 58 patients treated with the proposed 1mg/kg q2w during additional 18 months.

5.6. Total patient exposure in the target indication

73 patients

5.7. Study outlines

5.7.1. Dose-finding

Title of study:FB9702-01	
Investigators (Study center):	
Studied period: Not available	**Phase of development:** Phase I/II
Objectives **Primary:** effect on glycosphingolipids accumulation in the vascular endothelium in different organs **Secondary:** safety and PK	
Design: Phase I/II open-label, dose-finding, single centre study studying 3 different product doses, with two different posologies for 2 of the doses. Each of the 5 different schedules was given to 3 subjects for a total of 5 infusions each. Plasma samples were taken at and after every infusion, and tissue (skin and liver) biopsy samples were taken at baseline and following the fifth infusion. Heart and kidney biopsies were considered as optional. Dose finding was based on effects on plasma GL-3 levels.	
Number of patients, by arm: 3 patients per treatment arm **Intended sample size:** 15 patients **Populations for analysis:** 15 patients	
Main inclusion criteria: Patients were exclusively male and aged 16 and older with a confirmed diagnosis of Fabry disease and with largely unaffected kidneys, as determined by clinical and laboratory kidney function parameters	
Main Exclusion criteria:	
Test product, dose and mode of administration, duration: Agalsidase beta in one of five treatment regimens (3 patients per group): either 0.3 mg/kg, 1.0 mg/kg or 3.0 mg/kg administered once every 14 days, or 1.0 mg/kg or 3.0 mg/kg administered once every 48 hours, for a total of five infusions	
Reference therapy, dose and mode of administration, duration: Groups were compared for dose response. No reference was used.	
Criteria for evaluation (Definition, timing of assessments): **Main efficacy assessment:** concentrations of GL-3 in plasma and urine Also, biochemical and histological examinations of pre- and post-treatment biopsy samples of several organs. Secondary variables: changes in physiological function and clinical parameters such as pain and quality of life were employed. **Safety:** Adverse events **Pharmacokinetic data:** Description of main PK parameters	

Main results (only actual scenario)

Plasma concentrations of GL-3 were cleared rapidly and in a dose-dependent way in the 14 days dosing schedule.

Patients showed a reduction in plasma GL-3 concentrations following the second infusion in the 1 mg/kg and 3 mg/kg dose groups, whereas in the 0.3 mg/kg dose group GL-3 did not substantially decrease until the third or fourth infusion.

Patients receiving r-hαGAL every 48 hours showed a response from the fourth infusion - after 8 days of treatment.

All three doses were effective in clearing GL-3 from endothelial vasculature, with the lower dose (0.3 mg/kg) cohort demonstrating less consistent clearance.

In the kidney samples of the Phase I/II study, there was also a general trend for reduction of GL-3 in glomerular mesangial cells and interstitial cells as well as in cells from the tubules and collecting ducts, which held the greatest accumulation of GL-3.

In cardiac tissue, the GL-3 clearance occurred predominantly from the endothelium and to a much lesser extent in cardiomyocytes.

In the liver, histology samples demonstrated significant GL-3 clearance in endothelial cells and Kupffer cells.

Table 1. Trial FB9702-01: Demographics and baseline characteristics

	0.3/14-day *n=3*	1.0/14-day *n=3*	3.0/14-day *n=3*	1.0/48-hr *n=3*	3.0/48-hr *n=3*
Age (yr.) mean ± std error Series	41.0 ± 3.0 35, 44, 44	33.7 ± 3.4 27, 36, 38	34.7 ± 1.5 32, 37, 35	27.0 ± 5.5 37, 26, 18	35.7 ± 4.7 45, 30, 32
Weight (kg) mean ± std error Series	64.7 ± 4.1 66, 57, 71	73.6 ± 9.1 88, 57, 76	69.1 ± 7.4 56, 82, 69	69.8 ± 2.4 73, 65, 72	78.0 ± 8.0 74, 67, 93
Gender (n) male	3	3	3	3	3
Race (n) White Black Hispanic	 3 0 0	 1 0 2	 2 0 1	 2 0 1	 3 0 0
Serum creatinine Mean ± std error Series	1.3 ± 0.4 0.6, 1.3, 2.0	1.2 ± 0.1 1.4, 1.0, 1.1	1.4 ± 0.2 1.5, 1.6, 1.1	1.0 ± 0.3 1.6, 0.8, 0.6	1.6 ± 0.2 1.9, 1.7, 1.1
Plasma α-gal	BDL*	BDL*	BDL*	BDL*	BDL*
Prestudy plasma GL-3 (ng/ml) mean ± std Series	22.0 ± 3.1 16.6, 27.4, 22.1	15.2 ± 4.3 16.7, 21.7, 7.2	29.5 ± 10.7 48.6, 28.2, 11.6	20.0 ± 3.2 20.3, 14.3, 25.5	3.3 ± 1.6 6.2, 2.6**, 1.0**

*below 55 ng/μl
**below the inclusionary limit of ≥5 ng/ml

Plasma GL-3 levels

Levels of plasma GL-3 were reduced on average in all groups by the last infusion (Table 8). Levels of GL-3 in both the 1 and 3 mg/kg dose groups in the every-14-day infusion regimen appeared to drop to end-of-treatment levels by infusion 2. The 3mg/kg, every-48-hour infusion group had a very low baseline GL-3 level (2 subjects in the cohort were enrolled with levels below the inclusion criterion) and did not contribute to the understanding of dose and clearance.

Table 8. Trial FB9702-01: Plasma GL-3 levels (ng/ml)

Dose Group *n=3 each*	Pre-infusion 1 Mean (range)	Pre-infusion 2* Mean (range)	Pre-infusion 5** Mean (range)
0.3/14 day	18.2 (15.8 - 20.4)	11.9 (10.8 -13.8)	3.0 (0.9 - 5.0)
1.0/14 day	15.7 (12.5 - 20.3)	1.8 (0.0 – 5.0)	3.0 (0.0 - 8.6)
3.0/14 day	34.1 (20.3 - 53.9)	1.2 (0.6 -1.7)	0.8 (0.0 - 1.6)
1.0/48 hour	13.2 (3.0 – 23.5)	39.9 (16.2 – 80.0)	9.8 (4.6 – 13.0)
3.0/48 hour	4.3 (2.0 – 7.4)	3.8 (0.0 -11.4)	0.0 (0 - 0)

*day 14 for the every-14 day group; day 2 for the every-48 hour group
**day 56 for the every-14 day group; day 8 for the every-48 hour group

5.7.2. Pivotal study

Title of study: AGAL-1-002-98

Investigators (Study center):

Studied period: 1 year
First Patient Enrolled 14 March 1999
Last Patient Completed 04 February 2000

Phase of development:
Phase III

Objectives

Primary:

The primary objective of the study was to evaluate the safety and efficacy of recombinant human α-galactosidase (r-hα GAL) compared to placebo for the treatment of patients with Fabry disease.

Secondary: assessment of the effectiveness of r-hα GAL compared to placebo based on changes from Baseline to Visit 11 (Week 20) in

- McGill Pain Questionnaire (short form)
- composite score of GL-3 inclusions in the capillary endothelium (vasculature) of the kidney, skin, and heart.
- composite score of GL-3 levels, as measured by Enzyme Linked Immunosorbant Assay (ELISA) in kidney tissue and urine

Design:

This was a multinational, multicenter, placebo-controlled, double-blind, randomised study of patients with a current diagnosis of Fabry disease who had no prior treatment with r-hα GAL. Patients received approximately 1.0 mg/kg (0.9 to 1.1 mg/kg) of r-hα GAL or placebo every 2 weeks for 20 weeks (11 Patient Visits) for a total of 11 infusions of study medication. Twenty-eight additional days were allowed for some of the final safety and efficacy procedures associated with Visit 11 (Week 20). Total duration after the first infusion was up to 168 days.

Number of patients, by arm: 29 each

Intended sample size: 60 patients planned for enrollment; detailed explanation on the sample size calculation is not available, thus it is assumed that the sample is determined by feasibility and convenience.

Populations for analysis: Efficacy analyses for the primary endpoint were performed on the "Intent-To-Treat", the "As Treated," and the "Per Protocol" populations. There were 58 treated patients

Main inclusion criteria:

Patients who met all of the following inclusion criteria were eligible to participate in the study:

- Patients were > 16 years old and had a current diagnosis of Fabry disease with no prior treatment with r-hα GAL.
- Patients had documented plasma α-galactosidase (αGAL) activity of < 1.5 nmol/hr/mL or a documented leukocyte αGAL activity of < 4.0 nmol/hr/mg.
- Patients had a clinical presentation consistent with Fabry disease.
- Patients had to be able to comply with the clinical protocol, which required extensive clinical evaluations and completion of questionnaires.
- Female patients of childbearing potential had a negative pregnancy test (urine β -hCG) prior to dosing at each study visit. In addition, all female patients of childbearing potential were required to use a medically accepted method of contraception throughout the study.

Main Exclusion criteria:

- Patients with kidney failure or renal insufficiency (defined as serum creatinine >2.2 mg/dl) or those who had undergone kidney transplantation or were on dialysis were not included.

Test product, dose and mode of administration, duration:

r-hαGAL was supplied in 20-mL vials (35 mg/vial) as a lyophilized preparation. Each vial of r-hαGAL was reconstituted with 7.2 mL of sterile water for injection. The appropriate amount of reconstituted r-hαGAL was further diluted with a 0.9% sodium chloride solution to a final total volume of 500 mL.

Patients received approximately 1.0 mg/kg (0.9 to 1.1 mg/kg) of r-hαGAL intravenously every 2 weeks, for a total of 11 infusions. Patients were to receive their intravenous infusion at a rate of no more than 0.25 mg/min over approximately 6 hours. Adjustments that allowed for longer infusion rates were made in those patients who experienced suspected hypersensitivity reactions associated with the study medication.

Reference therapy, dose and mode of administration, duration:

Placebo: The placebo was a lyophilized preparation of mannitol with a phosphate buffer in vials identical to those used for r- hαGAL. Placebo vials were reconstituted with 7.2 mL of sterile water for injection and diluted with a 0.9% sodium chloride solution to a final total volume of 500 mL.

Criteria for evaluation (Definition, timing of assessments):

Efficacy assessments:

- The primary endpoint analysis was based on the number of patients who had a consensus (> 2 of 3 pathologists) score of zero for GL-3 accumulation from the capillary endothelium of the kidney after dosing with randomised study medication for 20 weeks. Three pathologists who measured the GL-3 inclusions in the capillary endothelium by semiquantitative light microscopy using a scale of 0-3 evaluated the tissue deposits. The reading was blinded for the treatment allocation and the sequence of the biopsies.

Secondary variables:

- Change from baseline in the composite score of GL-3 inclusions in the capillary endothelium of the kidney, skin and heart (assessed by LM);
- Change from baseline in composite score of kidney tissue and urinary GL-3 levels (measured by ELISA);
- Reduction in pain as assessed by the Short Form McGill Pain Questionnaire

Tertiary endpoints:

- Quality of Life (SF-36), change of glomerular filtration rate, neuropathy impairment, and autonomic function status.

Pharmacokinetic data / Pharmacodynamic data / other:

pharmacokinetic (PK) profile of r-hαGA and enzyme uptake in leukocytes in a subgroup (Europe)

Safety:

- Incidence of AEs and changes in vital signs, ECGs, ECHOs, physical examinations, and clinical laboratory safety parameters.

Statistical methods

Analysis of efficacy: Chi-square test for primary analysis of the primary endpoint (number of patients with a LM consensus score of the capillary endothelium of the kidney equal to 0 divided by the total number of patients with a score greater than 0 – *thus an odds, not a proportion?*) at 20 weeks. Also an ANOVA was used to test mean change scores between treatments and study centers. In addition, subgroup analyses based on age, ethnicity, and study site were evaluated to determine the effect of r-hαGAL on kidney LM score at week 20.

Secondary endpoints were analysed and presented similarly to the primary efficacy endpoint. These endpoints were analysed for the "Intent-To-Treat" and the "As Treated" populations.

Pharmacokinetics plasma concentration-time data and leukocyte uptake data were directly across 3 visits.

Heterogeneity was studied through hypothesis testing to determine if there were any statistically significant differences in demographics, history of Fabry disease, or medical/surgical history between the 2 treatment groups.

Missing data handling: The EPAR did not include details on missing data handling, besides the definitions of the subsets for analysis.

Tolerability and acceptability: Descriptive statistics only

Main results (only actual scenario)

In ITT; 18/29 (62%) patients in active vs 2/29 (7%) patients in the placebo group showed a final majority score of zero in kidney samples at 20 weeks (p<0.001), and similarly in As-Treated analyses. In the As-Treated population, no patients in the placebo group demonstrated a final majority score of zero, as the two patients randomised to placebo actually had received r-haGAL. Thus, 20 of the 29 (69%) r-haGAL-treated patients had a final majority score of zero as compared to none in the placebo-treated group.

Secondary endpoints:

- Reductions in GL-3 inclusions by LM were observed in the vascular endothelium of the kidney, skin and heart tissue using a combined severity score ranging from 0 to 3 for each organ. A score reduction of 86% was found in the r-haGAL treated patients compared to no change in the placebo patients (p<0.001).
- Consistent results in each organ (67% score reduction in heart and 100% score reduction in skin for r-haGAL group, both p<0.001).
- Change in urinary and kidney GL-3 as measured by ELISA: median change in urinary GL-3 of -23% vs +43% in placebo (p=0.005).
- Change in GL-3 levels in the kidney -34% vs -6% with placebo (NS)
- Normalization of plasma GL-3 in the r-haGAL treated patients, vs 17% decrease in placebo (p<0.001). Almost all patients receiving active had GL-3 values below the detection level of 1.2 ng/µl.
- Baseline values for McGill Short Form Pain Questionnaire were low (i.e., low to moderate pain). There was a statistically significant difference in the change from Baseline between the treatment groups; however, there was no difference between the 2 treatment groups at Visit 11 (Week 20).

Tertiary endpoints:

-No statistically significant differences between the treatment groups were observed in QoL

-No significant study site effect was noted and immunoglobulin G (IgG) antibody formation had no effect on efficacy results.

-The mean PK parameters of r-haGAL at Baseline (Infusion 1) show that total clearance (CL) averaged 1.75 ± 0.77 mL/min/kg, volume of distribution (Vz) averaged 0.23 ± 0.14 L/kg and volume of distribution,at steady-state (Vss) averaged 0.12 ± 0.08 L/kg. The mean elimination half-life (t½) was 88.6 ± 20.2 min and mean residence time was 66.4 ± 14.1 min. The PK parameters, area under the curve (AUC∞) and CL for r-haGAL following repeat administration were different at Infusion 7 and appeared to be associated with the formation of IgG antibodies to r-haGAL. The observed changes in PK did not appear to affect the efficacy outcomes. Cellular uptake of the enzyme, as measured in leukocytes was not reduced, but rather appeared to increase with repeat infusions.

Safety Results

There were no deaths, and no patients discontinued from the study because of AEs. The occurrence of SAEs was similar between the 2 treatment groups (5 patients experienced SAEs in each treatment group), and no SAE was considered related to study treatment.

All patients in each treatment group reported at least 1 AE during participation in the study. A statistically significant difference was observed for 3 AEs that were reported more frequently in patients treated with r-haGAL compared to patients treated with placebo. These AEs included rigors (52% versus 14%; p=0.004), fever (48% versus 17%; p=0.024), and skeletal pain (21% versus 0%; p=0.023).

5.8. Additional information for modelling

Proportion of patients in the agalsidase beta group achieved a final kidney majority score of 0 at week 20:

- ITT: 18/29 (62%) patients in the r-haGAL treatment group compared to 2/29 (7%) patients in the placebo group showed a final majority score of zero. (p<0.001)
- In the As-Treated population: 20 of the 29 (69%) r-haGAL-treated patients compared to none in the placebo-treated group. No patients in the placebo group demonstrated a final majority score of zero, as the two patients randomised to placebo actually had received r-haGAL.

Table 18. Trial AGAL-1-002-98: Primary endpoint results*

End of trial score	Placebo *n=29*	r-haGal *n=29*
Zero	0 (0%)	20 (69%)
Non-zero	29 (100%)	9 (31%)
Odds ratio (C.I.)	0.008 (0.00, 0.14)	
p-value	<0.001	

*as-treated population, considered equivalent to intent-to-treat

Tertiary endpoints:

- When analysing 6-month interim data from the ongoing phase III open-label extension study, results indicate that the renal function remained stable during the first year of the treatment thus suggesting a clinical effect.
- Glomerular filtration rate, planned for all subjects. The analysis presented by Genzyme (see Table 55, which combines data from AGAL-1-002-98 and its extension, AGAL-005-99) must be interpreted with caution: As a result of difficulties in obtaining inulin after baseline determinations were made, some subjects did not have end-of-treatment inulin tests as planned; for these subjects, GFR was estimated from creatinine clearance by multiplication by 0.77. In addition, baseline glomerular filtration rate was not balanced between the treatment groups, and end-of-trial calculations were performed in a subset of the trial population. GFR improved for the placebo as well as the r-haGal-treated group, which is a medically implausible result.

Table 55. GFR (mean ± st. dev) in AGAL-1-002-98 and at 6 months of AGAL-005-99

Trial	Visit	Statistic	Treatment group	
			placebo	r-hαGal
AGAL-11-002-98	Baseline	*N*	*28*	*29*
		Mean	97 ± 35	82 ± 22
	visit 11	*N*	*23*	*21*
		Mean	108 ± 39	93 ± 34
			placebo/ r-hαGal	r-hαGal/ r-hαGal
AGAL-005-99	6-month	*N*	*26*	*23*
		Mean	117 ± 41	82 ± 30

Table 56. GFR (mean ± std. error) by inulin technique only in AGAL-1-002-98 and at 6 months of AGAL-005-99 *

Trial	Visit	Treatment group*	
AGAL-1-002-98		Placebo *n=19*	r-hαGal *n=17*
	Baseline	96 ± 7	75 ± 6
	Visit 11	101 ± 8	94 ± 9
AGAL-1-002-98 and AGAL-005-99		Placebo *n=17*	r-hαGal *n=15*
	Baseline	96 ± 8	78 ± 6
	Visit 11	102 ± 9	93 ± 9
	6-month	116 ± 6	79 ± 8

*as-treated group, considered the same as intent-to-treat

Besides, it seems from previous data that patients with more advanced disease may be more sensitive to treatment effects, so a more sensitive population for clinical trials.

5.8. Uncertainties/weaknesses identified

The demonstration of efficacy was based on PD markers (reduction of sphingolipids in the target organs) with a complete absence of clinical endpoints, i.e. symptoms, function, etc. Although on a theoretical basis this may precede clinical improvement or a stabilisation of the clinical condition, none of the clinical parameters investigated as SEP did reach statistically significant improvement. Thus, the treatment is assessed in short term but assumed to provide a long term benefit that, in the intended population, would not be clinically evident until a long follow-up period had passed. From other information on the disease and trials with similar treatments, it can be derived that patients with more advanced disease may be more responsive to treatment, so that clinical changes may be quantified better in patients with advanced disease.

Therefore, at the time of decision making for marketing authorisation, there were uncertainties on how changes in sphingolipids may later translate into a clinically relevant effect, and inference on the potential benefit of the product was assumed to derive from the hypothesis and physiopathology.

Actually, adverse events (rigors, fever and skeletal pain) were more frequent with active treatment than with placebo, so that in clinical terms and according to available data, the effect could even be deleterious.

Such uncertainties were acknowledged, so that the authorisation was issued *under exceptional circumstances* and a number of post-marketing commitments were requested in order to collect additional long- term efficacy and safety data. The applicant committed to complete a program of clinical studies post-authorisation, the results of which shall form the basis of the annual reassessment of the benefit/risk profile.

6. Alternative development plans

6.1. Option 1: alternative development: RCT, placebo-controlled over SOC with sequential analysis for small populations

This option was selected based on the results of the systematic exercise of applicability of methods.

6.1.1. Safety and tolerability

Same as main development

6.1.2. Proof of activity and dose finding

Same as main development

6.1.3. Pivotal study

Group sequential designs for small populations could be applied if the PEP were replaced by a continuous endpoint, i.e. based on renal changes (CrCl). Given the concerns on the sensitivity of the trial to show an effect in CrCL, enrichment approaches might be required. The study will include interim analysis with stopping rules for both futility and superiority, for one or all arms.

Title of study: ALTERNATIVE AGAL-1-002-98 (simulated Option 1: Sequential design for small populations)
Investigators (Study center):
Studied period: 1 year from FPI to LPO, based on previous trials
Objectives: The primary objective of the study will be to evaluate the safety and efficacy of recombinant human α-galactosidase (r-hα GAL) compared to placebo for the treatment of patients with Fabry disease. **Secondary:** include assessment of the efficacy of r-hα GAL compared to placebo based on changes from Baseline to Visit 11 (Week 20) of at least one of the following variables: in renal function (creatinine clearance, 24h proteinuria), quality of life as assessed through SF36, and the proportion of patients with score 0 in the composite score of GL-3 inclusions in the capillary endothelium (vasculature) of the kidney, skin, and heart. Changes in the McGill Pain Questionnaire (short form) will be secondary, but it is unexpected that differences may be detected with the anticipated sample size and time of follow-up. The change from Baseline to Visit 11 (Week 20) in composite score of GL-3 levels, as measured by Enzyme Linked Immunosorbant Assay (ELISA) in kidney tissue and urine, will be also a secondary objective.
Design: Multinational, multicenter, placebo-controlled, double-blind, randomised study of patients with a current diagnosis of Fabry disease who have had no prior treatment with r-hα GAL. Patients will receive approximately 1.0 mg/kg (0.9 to 1.1 mg/kg) of r-hα GAL or placebo every 2 weeks for 20 weeks (11 Patient Visits) for a total of 11 infusions of study medication. Twenty- eight additional days will be allowed for some of the final safety and efficacy procedures associated with Visit 11 (Week 20). Therefore, the total duration that a patient will be involved in the study after the first infusion will be up to 168 days. A group sequential analysis for small sample sizes will be applied in an enriched population with pre-existing renal involvement/impairment.
Number of patients, by arm: Considering the differences observed in previous studies for the main variable, the sample size required to detect significant differences in creatinine clearance of 18 mL/min and SD of 21 mL/min after 6 months therapy would be about 24 subjects per group (48 in total, see table below with post-hoc calculations). Besides, the

resulting sample size would allow to detect significant differences also in some histopathology secondary assessments.
Since the average sample size saving that may be obtained in sequential designs is anticipated to be 30%, the expected sample size would be roughly 16 patients per arm (32 in total), thus allowing for early interruption in case of conclusive results. It is anticipated that the sample size may be reached in 1 year or less, based on the time required to recruit 58 patients in a previous study, with 6 months follow-up for assessment of the main end-point.

	Sample size per group	Differences Active vs Placebo*	SD pooled	Effect Size
Fabrazyme				
Primary Outcome: Grade 0 at week 20	8-9	-69%		
Secondary outcomes:				
sensory pain score	7715	0.3	6.65	0.045
affective pain score	520	0.4	2.3	0.174
total pain score	1774	0.8	8.5	0.094
visual analog scale score	884	-0.2	1.5	0.133
present pain intensity	639	0.4	2.55	0.157
Replagal				
Creatinine Clearance; 6 Month Data. TKT003	24	-18.1	21.653	0.836
Glomerular Filtration Rate; 6 Month Data				
TKT003	48	-11	18.93	0.581
TKT010	1749	1.2	12.66	0.095
Standard Renal Histopathology: Effects of Replagal				
Normal Glomeruli	11	-0.241	0.19	1.268
Fraction of Glomeruli with Mesangial Widening	9	-0.29	0.2	1.45
The Effects of Replagal on Cardiac Disease TKT005				
Cardica GB3	47	-0.18	0.306	0.588
Left Ventricular Mass by MRI	10	-33.3	23.68	1.406
Left Ventricular Mass by Echo	411	-12.5	63.86	0.196

Populations for analysis: Since it is difficult to define specific subsets for the sequential analysis, the population that would be included in the main analysis would be the randomised population. Sensibility analyses would be run thereafter for primary analysis, and as main analysis for secondary variables, on an ITT set (all subjects who were consented to participate in the protocol), and also sensitivity analyses will be done on the "as treated" and the "Per Protocol" analysis sets.

Main inclusion criteria:

- Patients > 16 years old and have a current diagnosis of Fabry disease with no prior treatment with r-hα GAL.
- Patients have documented plasma α-galactosidase (αGAL) activity of < 1.5 nmol/hr/mL or a documented leukocyte αGAL activity of < 4.0 nmol/hr/mg.
- Patients have a clinical presentation consistent with Fabry disease.
- Patients have to be able to comply with the clinical protocol, which required extensive clinical evaluations and completion of questionnaires.
- Female patients of childbearing potential have a negative pregnancy test (urine β -hCG) prior to dosing at each study visit. In addition, all female patients of childbearing potential will be required to use a medically accepted method of contraception throughout the study.
- Patients with symptoms (neuropathic pain) and/or functional impairment in kidney (such as creatinine above 2.2 mg/dL, reduced glomerular filtration below 55 mL/min per 1.73 m^2 or proteinuria) and/or increased thickness of the ventricular septum (+/- cardiac/SNP).

Main Exclusion criteria:

Test product, dose and mode of administration, duration:
Idem as pivotal

Reference therapy, dose and mode of administration, duration:
Placebo: The placebo will be a lyophilized preparation of mannitol with a phosphate buffer in vials identical to those used for r- hαGAL. Placebo vials will be reconstituted with 7.2 mL of

sterile water for injection and further diluted with a 0.9% sodium chloride solution to a final total volume of 500 mL.

Criteria for evaluation (Definition, timing of assessments):
Efficacy assessments:
Primary: Changes from baseline in renal function (creatinine clearance or 24h proteinuria),
Secondary:

- Change from baseline in the composite score of GL-3 inclusions in the capillary endothelium (vasculature) of biopsies from the kidney, skin, and heart (assessed by LM). Reduction of GL-3 accumulation from the capillary endothelium of the kidney will be assessed by 3 pathologists who will measure the GL-3 inclusions in the capillary endothelium by semiquantitative light microscopy using a scale of 0-3 for assessment of tissue deposits, in blind conditions with respect to the treatment identity and sequence of the biopsies.
- The number of patients who had a consensus (> 2 of 3 pathologists) score of zero at 20 weeks.
- Change in composite score of kidney tissue and urinary GL-3 levels (measured by ELISA);
- Reduction in pain as assessed by the Short Form McGill Pain Questionnaire
- Changes from baseline to Quality of life as assessed through SF36,

Tertiary endpoints:

- Neuropathy impairment, and autonomic function status, and others.

Safety:

- Incidence of AEs and changes in vital signs, ECGs, ECHOs, physical examinations, and clinical laboratory safety parameters.

Pharmacokinetic data / Pharmacodynamic data / other:

- Pharmacokinetic (PK) profile of r-hαGAL will be assessed in a selection of sites. Enzyme uptake in leukocytes will also be evaluated in these patients.

Statistical methods
Analysis of efficacy:
Efficacy analyses for the primary endpoint will be performed on the "Intent-To-Treat", the "As Treated," and the "Per Protocol" populations. A group sequential analysis for small sample sizes will be applied in an enriched population with pre-existing renal involvement/impairment. Inspections will be planned after 65% and 75% of patients will complete week 20 of follow-up. An ANOVA test will be done for comparing changes from baseline in the creatinine change. O'Brian-Fleming adjustment for multiplicity will be applied to account for multiple inspections. Stopping rules for both futility and superiority will be determined by the result of the O'Brian-Fleming adjustment to account for a maximum level of one-sided type 1 error of 5% at the study level. Secondary analyses will be done using Fisher's exact test or ANOVA test according to the type of variable.
Missing data handling: Measures to minimise loss of information will be implemented, including assessments post-study treatment interruption. Sensitivity analyses will take into account post drop-out measurement and will consider the potential role of the treatment toxicity as potential source of missing data.
Tolerability and acceptability: Descriptive analyses of treatment emergent and related (at least possible) adverse events by seriousness, severity, and expectedness will be displayed by treatment group for the "as treated" population.

6.2. Option 2: Study with multi-arm multi-stage trial with a simultaneous stopping rule

This option was selected based on the results of the systematic exercise of applicability of methods.

6.2.1. Safety and tolerability

Same as main development

6.2.2. Dose finding and pivotal efficacy

Two studies: first, same as main development for proof of concept and exploration of doses, including 15 subjects to test 3 different doses, two of them with two different schedules, total of 5 treatments, given to 3 patients each. After preliminary dose selection, a dose selection with pivotal confirmation could follow.

Title of study: FB9702-01 (dose-finding Study with multi-arm multi-stage trial with a simultaneous stopping rule)	
Investigators (Study center):	
Studied period: 18 months to recruit 72 patients, according to previous studies	**Phase of development:** Phase II/III
Objectives **Primary:** effect on glycosphingolipids accumulation in the vascular endothelium in different organs **Secondary:** safety and PK	
Design: Phase II/III double-blind, dose-finding, multi-centre study testing two doses of Agalsidase beta vs placebo up to 20 weeks..	
Number of patients, by arm: 24 patients per treatment arm **Intended sample size:** 72 patients **Populations for analysis:** Intention to treat, as treated and per protocol subsets.	
Main inclusion criteria: • Patients > 16 years old and have a current diagnosis of Fabry disease with no prior treatment with r-hα GAL. • Patients have documented plasma α-galactosidase (αGAL) activity of < 1.5 nmol/hr/mL or a documented leukocyte αGAL activity of < 4.0 nmol/hr/mg. • Patients have a clinical presentation consistent with Fabry disease. • Patients have to be able to comply with the clinical protocol, which required extensive clinical evaluations and completion of questionnaires. • Female patients of childbearing potential have a negative pregnancy test (urine β -hCG) prior to dosing at each study visit. In addition, all female patients of childbearing potential will be required to use a medically accepted method of contraception throughout the study. • Patients with symptoms (neuropathic pain) and/or functional impairment in kidney (such as creatinine above 2.2 mg/dL, reduced glomerular filtration below 55 mL/min per 1.73 m^2 or proteinuria) and/or increased thickness of the ventricular septum (+/- cardiac/SNP).	
Main Exclusion criteria:	
Test product, dose and mode of administration, duration: Agalsidase beta in two treatment regimens: 1.0 mg/kg or 3.0 mg/kg administered once every 14 days, up to 20 weeks. The medication will be prepared by a third party not involved in the trial.	
Reference therapy, dose and mode of administration, duration:	

Identical matching placebo to both regimes. The medication will be prepared by a third party not involved in the trial.
Criteria for evaluation (Definition, timing of assessments): Renal function will be assessed through creatinine clearance. Plasma samples and tissue (skin and liver) biopsy samples will be obtained at baseline and end of study treatment to measure concentration of GL-3 and deposition in tissues. **Main efficacy assessment:** Change from baseline in creatinine clearance. **Secondary:** changes from Baseline to Visit 11 (Week 20) of 24h proteinuria, quality of life as assessed through SF36, and the proportion of patients with score 0 in the composite score of GL-3 inclusions in the capillary endothelium (vasculature) of the kidney, skin, and heart. Changes in the McGill Pain Questionnaire (short form) will be secondary, but it is unexpected that differences may be detected with the anticipated sample size and time of follow-up. The change from Baseline to Visit 11 (Week 20) in composite score of GL-3 levels, as measured by Enzyme Linked Immunosorbant Assay (ELISA) in kidney tissue and urine, will be also a secondary objective. **Safety:** Descriptive analyses of treatment emergent and related (at least possible) adverse events by seriousness, severity, and expectedness will be displayed by treatment group for the "as treated" population. **Pharmacokinetic data:** Descriptive PK data, immunogenicity
Statistical methods **Analysis of efficacy:** A group sequential analysis will be done with inspections planned after 65% and 75% of patients will complete week 20 of follow-up. An ANOVA testing will be applied for comparing the change from baseline in creatinine clearance between groups. Simultaneous stopping rules will be defined based on first conclusion of superiority for one of the arms vs others, according to the method of Urach et al, preserving cumulatively a maximum unwanted error. If the trial will be stopped early, treatment efficacy testing will be performed based on the interim data accumulated. **Missing data handling:** According to the experience from previous trials the amount of missing data will be very low, however, measures to minimise loss of information will be implemented, including assessments post-study treatment interruption. Sensitivity analyses will take into account post drop-out measurement and will consider the potential role of the treatment toxicity as potential source of missing data. **Tolerability and acceptability:** Descriptive analyses of treatment emergent and related (at least possible) adverse events by seriousness, severity, and expectedness will be displayed by treatment group for the "as treated" population.

6.3. Option 3: Fallback test for co-primary endpoints + enrichment

This option was selected based on the clustering of conditions, reviewing the clinical characteristics of the intended therapeutic indication.

6.3.1. Safety and tolerability

Same as main development

6.3.2. Proof of activity and dose finding

Equal to main development

6.3.3. Pivotal evidence

Because the disease may affect kidney function, may induce neuropathic pain as one of the key symptoms, has heterogeneity and difficulties to measure treatment consequences on health perception, and reasonable dynamic markers can be

measured through histology of kidney, skin and heart, there may be room to application of multiple end-points. Thus, multiple endpoints including clinical outcome measures (symptomatic changes, functional and QoL) and histological changes may provide a more convincing (clinically relevant) demonstration of efficacy.

The proposal thus may include methods based on multiple endpoints (i.e. like the fallback tests for co-primary endpoints and the optimal exact tests for multiple binary endpoints), coupled with enrichment methods, where patients with clinical symptoms or functional impairment are included.

Title of study: ALTERNATIVE AGAL-1-002-98 (simulated Option 3) : fallback test for co-primary endpoints + enrichment

Investigators (Study center):

Studied period: 1 year, based on previous studies

Objectives

Primary: to evaluate the efficacy of recombinant human α-galactosidase (r-hα GAL) compared to placebo for the treatment of patients with Fabry disease, based on changes from baseline to visit 11 (Week 20) of at least one of the following variables: renal function (creatinine clearance, 24h proteinuria), bodily pain domain of the SF36, and the proportion of patients with score 0 in the composite score of GL-3 inclusions in the capillary endothelium (vasculature) of the kidney, skin, and heart.

Secondary:

- Safety of r-hα GAL
- Change from baseline to visit 11 (Week 20) in composite score of GL-3 levels, as measured by Enzyme Linked Immunosorbant Assay (ELISA) in kidney tissue and urine.
- Changes in the McGill Pain Questionnaire (short form). However, it is unexpected that differences may be detected in MG-PQ with the anticipated sample size and time of follow-up.

Design: Multinational, multicenter, placebo-controlled, double-blind, randomised study of patients with a current diagnosis of Fabry disease with clinical expression and who have had no prior treatment with r-hα GAL. Patients will receive approximately 1.0 mg/kg (0.9 to 1.1 mg/kg) of r-hα GAL or placebo every 2 weeks for 20 weeks (11 Patient Visits) for a total of 11 infusions of study medication. The total duration that a patient will be involved in the study after the first infusion will be up to 168 days.

Sample size: The size of the study is based on patient availability and is thus marked by convenience and feasibility (60 patients, 30 per arm).

Post-hoc sample size calculations: Based on the observed active (Fabrazyme and Replagal) vs placebo differences and variability in previous studies, and assuming a two-sided 5% type I error and 80% statistical power, the following sample sizes would be required to detect statistical significance assuming similar size of effect as observed in previous trials.

	Sample size per group	Differences Active vs Placebo*	SD pooled	Effect Size
Fabrazyme				
Primary Outcome: Grade 0 at week 20	8-9	-69%		
Secondary outcomes:				
sensory pain score	7715	0.3	6.65	0.045
affective pain score	520	0.4	2.3	0.174
total pain score	1774	0.8	8.5	0.094
visual analog scale score	884	-0.2	1.5	0.133
present pain intensity	639	0.4	2.55	0.157
Replagal				
Creatinine Clearance; 6 Month Data. TKT003	24	-18.1	21.653	0.836
Glomerular Filtration Rate; 6 Month Data				
TKT003	48	-11	18.93	0.581
TKT010	1749	1.2	12.66	0.095
Standard Renal Histopathology: Effects of Replagal				
Normal Glomeruli	11	-0.241	0.19	1.268
Fraction of Glomeruli with Mesangial Widening	9	-0.29	0.2	1.45
The Effects of Replagal on Cardiac Disease TKT005				
Cardica GB3	47	-0.18	0.306	0.588
Left Ventricular Mass by MRI	10	-33.3	23.68	1.406
Left Ventricular Mass by Echo	411	-12.5	63.86	0.196

Main inclusion criteria:

- Patients > 16 years old and have a current diagnosis of Fabry disease with no prior treatment with r-hα GAL.
- Patients have documented plasma α-galactosidase (αGAL) activity of < 1.5 nmol/hr/mL or a documented leukocyte αGAL activity of < 4.0 nmol/hr/mg.
- Patients have a clinical presentation consistent with Fabry disease.
- Patients have to be able to comply with the clinical protocol, which required extensive clinical evaluations and completion of questionnaires.
- Female patients of childbearing potential have a negative pregnancy test (urine β -hCG) prior to dosing at each study visit. In addition, all female patients of childbearing potential will be required to use a medically accepted method of contraception throughout the study.

Also, a minimum degree of clinical expression will be required:

- Patients with symptoms (neuropathic pain) and/or functional impairment in kidney (such as creatinine above 2.2 mg/dL, reduced glomerular filtration below 55 mL/min per 1.73 m^2 or proteinuria) and/or increased thickness of the ventricular septum (+/- cardiac/SNP).

Main Exclusion criteria:

Test product, dose and mode of administration, duration:

r-hαGAL will be supplied in 20-mL vials (35 mg/vial) as a lyophilized preparation. Each vial of r-hαGAL will be reconstituted with 7.2 mL of sterile water for injection. The appropriate amount of reconstituted r-hαGAL will be further diluted with a 0.9% sodium chloride solution to a final total volume of 500 mL. Patients received approximately 1.0 mg/kg (0.9 to 1.1 mg/kg) of r-hαGAL intravenously every 2 weeks, for a total of 11 infusions. Patients will receive their intravenous infusion at a rate of no more than 0.25 mg/min over approximately 6 hours. Adjustments are allowed for longer infusion rates in those patients who experience suspected hypersensitivity reactions associated with the study medication.

Reference therapy, dose and mode of administration, duration:

Placebo: The placebo will be a lyophilized preparation of mannitol with a phosphate buffer in vials identical to those used for r- hαGAL. Placebo vials will be reconstituted with 7.2 mL of

sterile water for injection and further diluted with a 0.9% sodium chloride solution to a final total volume of 500 mL.

Efficacy assessments:

- The primary efficacy endpoint for this study will be efficacy of r-hα GAL compared to placebo based on the assessments at Visit 11 (Week 20) after dosing with randomised study medication for 20 weeks of the following variables, of which at least one will be significantly different from placebo:
  - Changes from baseline in renal function (creatinine clearance or 24h proteinuria), or
  - Changes from baseline in the bodily pain domain of the SF36 QoL Questionnaire, or
  - Change from baseline in the composite score of GL-3 inclusions in the capillary endothelium (vasculature) of biopsies from the kidney, skin, and heart(assessed by LM), or
  - The number of patients who had a consensus (> 2 of 3 pathologists) score of zero at 20 weeks.

  Reduction of GL-3 accumulation from the capillary endothelium of the kidney will be assessed by 3 pathologists who will measure the GL-3 inclusions in the capillary endothelium by semiquantitative light microscopy using a scale of 0-3 for assessment of tissue deposits, in blind conditions with respect to the treatment identity and sequence of the biopsies.

Secondary variables:

- The number of patients who had a consensus (> 2 of 3 pathologists) score of zero at 20 weeks.
- The change in composite score of kidney tissue and urinary GL-3 levels (measured by ELISA);
- Reduction in pain as assessed by the Short Form McGill Pain Questionnaire

Several tertiary endpoints:

- Neuropathy impairment, and autonomic function status, SF-36 and others.

Safety:

- Safety will be measured in terms of the incidence of AEs and changes in vital signs, ECGs, ECHOs, physical examinations, and clinical laboratory safety parameters.

Pharmacokinetic data / Pharmacodynamic data / other:

- Pharmacokinetic (PK) profile of r-hαGAL will be assessed in a selection of sites. Enzyme uptake in leukocytes will also be evaluated in these patients.

Statistical methods

Analysis of efficacy:

Efficacy analyses for the primary endpoint will be performed on the "Intent-To-Treat" (primary analysis), the "As Treated," and the "Per Protocol" populations. A mixed model for repeated measurements approach including the baseline measurement as a covariate will be fitted to test for a significant difference in 20-week mean change scores between treatments. The test will be applied to the four efficacy variables and Fallback tests strategy for co-primary endpoints will be applied.

The secondary endpoints will be analysed and presented similarly to the primary efficacy endpoint. Fisher exact tests will be applied to the analyses of binary parameters. These will be analysed for the "Intent-To-Treat" and the "As Treated" populations.

A PK analysis that directly compared the r-hαGAL plasma concentration-time data across 3 visits will be performed. A similar analysis will be performed across the same time points for the leukocyte uptake data.

Missing data handling: According to the experience from previous trials the amount of missing data will be very low; however, measures to minimise loss of information will be implemented, including assessments post-study treatment interruption. Sensitivity analyses will take into

account post drop-out measurement and will consider the potential role of the treatment toxicity as potential source of missing data.

Tolerability and acceptability: Descriptive analyses of treatment emergent and related (at least possible) adverse events by seriousness, severity, and expectedness will be displayed by treatment group for the "as treated" population.

7. Analysis of the practical, ethical and regulatory impact

7.3. Option 1: sequential design for small populations

Method assessed: Option 1: sequential design for small populations	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements	Yes	It is considered that, depending on the magnitude of the effect, group-sequential designs can achieve about 30% reduction of sample size, but may also increase study size in some circumstances.
• May shorten time to study completion	Yes	Theoretically yes, but it all depends on the ability to show a sizeable effect on renal function at 20 weeks, although length of recruitment as related to time of follow-up per patient does not leave much room for improvement.
• May ease recruitment	No (worsens)	More strict inclusion criteria, less eligible patients. Logistically more complex if new centers need to be opened during the conduct of the study.
Statistical assessment:		
• Improves internal validity	Similar	
• Increases stability of estimates	No	Probably not, in fact the decision based on the interim analyses may be subject to less stability in the estimates
• Increases sensibility to changes	Yes	Based on the use of a continuous endpoint, and enrichment of study population
• Compliant with predetermination	Yes	
• Consistency (discuss)	No (may worsen)	The ability to show an effect in secondary endpoints may be reduced if a reduced sample size or shorter term FU
• Robustness of method (discuss)	Similar	
• Protection against type I and II errors (discuss)	Similar	Sequential designs for small populations do adjust to avoid multiplicity issues.
Regulatory assessment:		
• Risk of bias and credibility	Controlled	Parallel double blind design, appropriate methods for analysis

Method assessed: Option 1: sequential design for small populations	Improves?	Comments
• External validity (discuss)	Yes	Enriched population may decrease external validity, but in this case, with such a good mechanistic rationale and with consistent results in substrate clearance from other tissues, not a major issue. Clinical end-points preferred over biomarkers.
• Therapeutic positioning and comparisons	No	Only compared to placebo
• Informative on relevance and clinical impact	Yes	Improved as PEP based on a clinical outcome
• Enough information on safety	Yes	Probably less data and shorter FU time, thus more uncertainty on the safety part.
• Suitable information for risk-benefit balance	Yes	It may reduce safety data due to sample size reduction in an already limited safety database
Ethical assessment:		
• May minimise risks	Yes	More advanced patients, but less numbers exposed
• May maximize access to treatment	No	Access to the therapeutic test (I e: entering the trial and having chances to receive active) is reduced, since only severe patients may participate.
• May minimise unnecessary exposure to ineffective treatments or placebo	Yes	Yes, as long as decisions can be taken earlier and thus the sample size can be reduced. However, there is no margin for improvement vs actual development considering the UR status of the condition, and the low number of patients actually included.
• Considers patient input	No	

Application of sequential designs for small sample sizes requires a change in the PEP to use a continuous endpoint (i.e. CrCL) and also enrichment of the study population to overcome the poor sensitivity of the study to changes in renal function. Sequential designs for small populations may reduce sample size and may shorten time to study completion. Nevertheless, this practical advantages are counterbalanced in this particular case by the fact that patients are already available, and at the time the interim analyses is performed, most might have already been enrolled and followed.

So, sequential design might delay access to treatment to less severe patients making the general ethical advantages of sequential designs, i.e. minimize exposure to placebo and to the experimental arm, of less value.

Therefore, the usual efficiency of sequential designs is of less relevance in this particular case. By contrary, this design may decrease external validity, the robustness of the evidence provided as it is based on an IA, and may also reduce the extent of an already limited safety database.

Furthermore, it might also increase logistical complexity and may delay study completion if new centers were to be opened during the conduct of the trial. For this reason, it is not considered the optimal approach.

7.2. Option 2 multi-arm multi-stage trial with a simultaneous stopping rule

Method assessed: Option 2: multi-arm multi-stage trial with a simultaneous stopping rule	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements	No	Increased because more arms are used
• May shorten time to study completion	Yes	Yes, theoretically, but in this particular case the actual development would not be shortened. Initial dose exploration study is still necessary, and three arms is longer than two arms pivotal. Besides, follow-up is 6 months and study duration 1 year, so that by half of study conduction (6m) all patients would have been randomised already, so there is no expected sample saving.
• May ease recruitment	No	Larger sample size, but options to receive active are bigger, so may stimulate participation
Statistical assessment:		
• Improves internal validity	Similar	
• Increases stability of estimates	No	Probably not, in fact the decision based on the interim analysis may be subject to less stability in the estimates
• Increases sensibility to changes	No	Contrast similar to conventional parallel
• Compliant with predetermination	Yes	Yes, interim pre-planned and methods for adjustment in place
• Consistency (discuss)	No	The ability to show an effect in secondary endpoints may be reduced if a reduced sample size or shorter term FU
• Robustness of method (discuss)	Yes	Based on pre-established criteria for dose-selection
• Protection against type I and II errors (discuss)	Similar	Methods for adjustment in place
Regulatory assessment:		
• Risk of bias and credibility	No	Parallel double blind design, appropriate methods for analysis
• External validity (discuss)	No	Larger size, but small numbers per group, yet inference has few supportive data
• Therapeutic positioning and comparisons	No	Compared only with placebo
• Informative on relevance and clinical impact	No	Improved as PEP based on a clinical outcome

Method assessed: Option 2: multi-arm multi-stage trial with a simultaneous stopping rule	Improves?	Comments
• Enough information on safety	Yes	Little impact on the overall development, but not efficient to increase exposure to the relevant dose level
• Suitable information for risk-benefit balance	No	Better information on dose selection, dose-response as an additional evidence of efficacy
Ethical assessment:		
• May minimise risks	No	No impact
• May maximise access to treatment	No	Randomsation rate to active:placebo 2:1, but half of active patients assigned to a dose different form the selected one
• May minimise unnecessary exposure to ineffective treatments or placebo	Yes	Yes, but little room for improvement because of quick recruitment as compared to patient follow-up
• Considers patient input	No	

In summary, option 2 may improve credibility in dose-selection, theoretically minimise exposure to ineffective treatments, and reduce time to completion, but given the little room for improvement this is not deemed a major added value, in particular since this approach does not address the relevant uncertainties identified in the actual development plan.

7.3. Option 3: fallback test for co-primary endpoints + enrichment

Method assessed: Option 3: fallback test for co-primary endpoints + enrichment	Improves?	Comments
Practical considerations:		
• May reduce sample size requirements	NO	Same sample size, determined by ultra-rarity
• May shorten time to study completion	NO	Same duration per patient, still too short to collect relevant clinical outcomes. Overall study duration may increase if difficult to find eligible patients with more severe disease
• May ease recruitment	NO (worsens)	More strict inclusion criteria, less eligible patients Same chances to receive placebo, less willingness to participate
Statistical assessment:		
• Improves internal validity	YES	
• Increases stability of estimates	NO	Depending on sample size, which is unmodified
• Increases sensibility to changes	YES	May reduce chances of failure for main endpoint with similar sample size, more chances to reach conclusive results.
• Compliant with predetermination	YES	

Method assessed: Option 3: fallback test for co-primary endpoints + enrichment	Improves?	Comments
• Consistency (discuss)	YES	Yes, proper assessment of the multidimensional nature of the disease
• Robustness of method (discuss)	YES	As demonstrated in the publication
• Protection against type I and II errors (discuss)	YES	Fallback method preserving from errors
Regulatory assessment:		
• Risk of bias and credibility	Controlled	Parallel double blind design, appropriate methods for analysis
• External validity (discuss)	YES	Enriched population may decrease external validity, but in this case, with such a good mechanistic rationale and with consistent results in substrate clearance from other tissues, not a major issue
• Therapeutic positioning and comparisons	YES	Improved, allowing comparison with trials in more severe populations
• Informative on relevance and clinical impact	YES	Improved substantially, since relevant variables may be conclusive from a confirmatory perspective
• Enough information on safety	YES	At least similar or higher, data in a more advanced (frail) set of patients
• Suitable information for risk-benefit balance	YES	Improved substantially
Ethical assessment:		
• May minimise risks	NO	More advanced (frail) set of patients may be more susceptible to adverse reactions
• May maximise access to treatment	NO	Access to the therapeutic test (I e: entering the trial and having chances to receive active) is reduced, since only severe patients may participate
• May minimise unnecessary exposure to ineffective treatments or placebo	NO	Same number of patients exposed to placebo for the same period of time. On top, less patients may have access to the active drug in the experimental setting because of strict inclusion criteria
• Considers patient input	YES	QoL as one of the primary co-endpoints

In summary, two main design modifications are introduced respect to the actual clinical development of Fabrazyme: an enriched design with more strict inclusion and exclusion criteria, aimed to select more severe patients who have more chances to be able to show treatment-related changes vs placebo, and the use of multiple co-primaries that may be each independently able to conclude confirmation of efficacy by application of a Fallback test for co-primary endpoints.

The advantages of this method are basically that the confirmatory efficiency is increased, as a result of both the likely increase of the effect size by selection of patients, and the reduction of risk of failing on the choice of the primary variable, by using a number of related variables with similar and complementary clinical relevance at the same level of confirmatory validity. Clinical consistency is also likely increased, thus potentially reducing uncertainties at the time of assessment of clinical benefit. The effect on the robustness of safety data is neutral or slightly favorable, since the population is anticipated to be more susceptible to adverse reactions, and is the target population where there are more causes for concern about risks from a clinical perspective.

Disadvantages may come from the ethical point of view, since the selection of subjects for an enriched design limits the number of subjects who may access the treatment within an experimental setting, and potential delays due to increased recruitment difficulties may lead to longer time to complete the pivotal evidence and thus later regulatory access to the new drug. As regards to the exposure to placebo, the method is neutral, although less patients may be eligible to participate in the trial because of strict inclusion criteria, as already explained. Considering that the condition has quite a slow progression, this may not seem a critical point with severe prognostic impact.

In conclusion, the overall balance of an enriched design is a reduction of uncertainty at the price of slower access to active treatment for mildly diseased patients. Using Fallback tests for co-primary endpoints is improving the trial at no substantial impact on other assessment parameters, and thus should be recommended.

8. Recommendations

The development of Fabrazyme in the treatment of Fabry´s disease is considered a representative model within the cluster of progressive multidimensional multi organ conditions. The pathophysiology of the disease is well known and thus the scientific rationale for the development of an ERT is strong. As for similar conditions, this is a chronic life-lasting disease with a highly variable clinical course, with impact in multiple system/organs, requiring multidimensional assessment and endpoints, relying often on subjective assessments from caregivers/patients on clinical or functional status and QoL.

Therefore, previous considerations on applicability of novel methods can reasonably be considered suitable options for conditions belonging to the cluster of chronic progressive conditions led by multiple organs/systems.

In particular, new methodologies aimed to study the multidimensional nature of the condition, like the Fallback tests for co-primary endpoints and the optimal exact tests for multiple binary endpoints, are highly recommended in order to generate a more complete and compelling evidence of efficacy and safety and to facilitate generalisation of the study results.

Parallel designs will be generally needed, due to progression of the condition and intersubject variability. Enrichment/stratification may be useful to control heterogeneity.

Previous information on the clinical course can be suitable for Bayesian approaches and planning of adaptions. Sample size adaptions and sequential designs, although applicable, are not considered increasing the efficiency if patients are already available for study entry, and since the use of placebo does not cast major ethical/practical concerns.

Annex 4. 6. Opsumit® (macitentan)

Clinical Development Plan Opsumit® (macitentan) Clinical development in pulmonary arterial hypertension

Version 1 22/08/2017

1 Index

2 Introduction

2.1 Background

2.1.1 Disease and currently available alternatives

Pulmonary Arterial Hypertension (PAH) is a chronic and progressive disease of the small pulmonary arteries that is characterised by vascular proliferation and remodelling. It results in increased pulmonary artery pressure and pulmonary vascular resistance and, ultimately, right ventricular heart failure and death. Although the pathogenesis of PAH is not completely understood, it likely involves an imbalance in the normal relationships between vasodilators and vasoconstrictors, growth inhibitors and mitogenic factors, and antithrombotic and pro-thrombotic determinants that are probably consequences of pulmonary endothelial cell dysfunction and/or injury.

PAH is defined by right-heart catheterisation showing a precapillary pulmonary hypertension (mean pulmonary artery pressure > 25 mmHg at rest or > 30 mmHg with exercise, with a pulmonary artery wedge pressure < 15 mmHg). There is a female-to-male preponderance (1.7:1), with patients most commonly presenting it in the third and fourth decade of life; although the age range is from infancy to greater than 60 years. The annual incidence of idiopathic pulmonary hypertension has been estimated within 1 or 2 cases per million individuals per year.

Current clinical classification of PAH comprises apparently heterogeneous conditions, which share comparable clinical and hemodynamic pictures and virtually identical pathologic changes of the lung microcirculation. PAH includes idiopathic (IPAH, formerly termed primary pulmonary hypertension) and familial forms (FPAH), PAH associated with various conditions (APAH), such as scleroderma and other connective tissue diseases (CTD), congenital heart defects with systemic-to-pulmonary shunts, portal hypertension, human immunodeficiency virus infection, exposure to drugs and toxins and other more rare settings: thyroid disorders, glycogen storage disease, Gaucher´s disease, hereditary heemorrhagic telangiectasia, hemoglobinopathies (Sickle disease especially), myelo-proliferative disorders, splenectomy, PAH associated with significant venous or capillary involvement and finally, persistent pulmonary hypertension of the newborn.

The functional classification is the measure of the limits imposed on a patient by a disease. There are two classification systems that are used interchangeably: The New York Heart Association (NYHA) functional classification and the World Health Organization (WHO) functional assessment classification.

Symptoms of PAH include dyspnoea (most commonly), fatigue, chest pain or discomfort, dizziness, syncope, near syncope, oedema, leg oedema, and palpitations. When the disease is advanced, the clinical manifestations include cyanosis, dyspnoea on exertion, haemoptysis, atypical chest pain or angina pectoris, syncope, heart failure, arrhythmias and cerebrovascular accidents.

At present, conventional treatment for patients with PAH includes calcium-channel blockers, anticoagulants, diuretics and oxygen. In addition, two phosphodiesterase-5 inhibitors

(sildenafil, tadalafil), two oral endothelin-receptor antagonists (ERA) (bosentan, ambrisentan), an intravenous prostacyclin (epoprostenol), an inhaled prostacylin (iloprost) and a subcutaneous prostacyclin (treprostinil) have also been licensed for the treatment of PAH in various European countries. Sildenafil, tadalafil and ambrisentan are indicated for patients with primary and CTD-associated PAH, while bosentan is indicated for patients with primary PAH, scleroderma-associated PAH and PAH associated to congenital systemic-to-pulmonary shunts and Eisenmenger's physiology. Iloprost is indicated only for patients with primary PAH. Four of these medicinal products are licensed for patients with NYHA FC II and III disease severity (sildenafil, tadalafil, bosentan and ambrisentan), whereas the remaining ones are licensed for patients with NYHA class III only.

2.1.2 Rationale for the development

Macitentan (ACT-064992) (N-[5-(4-Bromophenyl)-6-[2-[(5-bromo-2-pyrimidinyl)oxy]ethoxy]-4-pyrimidinyl]-N'-propylsulfamide) is an orally active, dual endothelin (ET) receptor antagonist. In vitro, macitentan selectively inhibits the binding of ET-1 to ETA and ETB receptors as well as the effects mediated by these receptors in functional assays.

2.2 Scope of development

The product is currently intended for the treatment of:

"Opsumit, as monotherapy or in combination, is indicated for the long-term treatment of pulmonary arterial hypertension (PAH) in adult patients of WHO Functional Class (FC) II to III. Efficacy has been shown in a PAH population including idiopathic and heritable PAH, PAH associated with connective tissue disorders, and PAH associated with corrected simple congenital heart disease."

The scope of development is from initial clinical data and decision to proceed to regulatory development to pivotal evidence achieved, and consistent with the Target Product Profile.

2.2.1 Target product profile

Indication	Long term treatment of Pulmonary arterial Hypertension (PAH) in adult patients with WHO functional Class II to III.
Route of administration	Oral
Pharmaceutical form	Film coated tablet 10 mg
Posology	10 mg/ once daily
Main target population	Adult patients with PAH WHO class II to III
Claims to be supported by the clinical development.	Macitentan reduces the risk of morbidity and mortality in adult patients with PAH Grade stage II to III

Regions where the product should be marketed:	Global (or EU + USA)
Regulatory agencies that will be involved	EMA FDA

3 General investigational plan

3.1 Objective (s) of the development

The objective of the development is to provide pivotal support to the application for marketing authorisation in the EU of the product, by generating:

- confirmatory evidence of (dose-response) AND/OR (superiority to placebo) in patients with PAH WHO FC II to III,
- an appropriate safety database, including enough information to permit a characterization of the safety profile of the product, and a risk-benefit assessment of the product at the time of assessment of the marketing authorisation application in the EU,
- any needed supporting information on the product to apply for a new drug authorisation in EU.

4 Assessment of applicability of methods

4.1 Representativeness of Opsumit within the cluster

Opsumit (macitentan) has been chosen as a potentially representative example of Staged Conditions within the ASTERIX clustering of medical conditions. It is a progressive condition with clearly defined clinical stages which relates with the prognosis of the patient. Although PAH is a fatal disease, and any treatments should be expected to modify survival, mortality is not usually found in clinical trials for products intended for PAH as a primary endpoint. Alternatively, other endpoints such as PAH related morbidity have been more often used, among them, the deterioration/improvement of functional class (WHO FC, based on clinical symptoms) and exercise capacity.

Some aspects, however, should be considered regarding this example that may differ from other conditions in the cluster:

- Other endothelin receptor antagonists have been developed previously, and the efficacy of targeting endothelin receptors has been demonstrated in several clinical trials and metaanalysis.
- A biomarker of the drug activity on the target receptor exists, thus PK-PD modelling is feasible.
- A guideline for the development of drugs for PAH was issued in 2008 (EMEA/CHMP/EWP/356954/2008).

4.2 Applicability of novel methodologies based on UMCU report

Applicable methodologies:

- Long-short term outcomes
- Sample size reassessment and hypothesis testing in adaptive survival trials
- Multi-arm group sequential designs with a simultaneous stopping rule
- GAS

Might be applicable:

- Sequential designs for small samples
- Bayesian sample size re-estimation using power priors
- Dynamic borrowing through empirical power priors
- Fallback tests for co-primary endpoints

Limited or non-applicability of the method

- Heterogeneity estimators
- Prior distributions for variance parameters in sparse-event MA
- Optimal exact tests for multiple binary endpoints
- Simultaneous inference for multiple marginal GEE models

Additional considerations:

Even if adaptive and sequential approaches are immediately applicable, recruitment could be faster than the time to endpoint depending on the population selected (more or less severe stages of the disease) and the endpoint chosen. However, due to the nature of the variable (time to event), the methodology for sample size reassessment in adaptive survival trials is applicable.

Depending on target population of the study, a short period to response assessment would allow adaptions and sequential approaches in general. Thus, a sequential design with arm dropping adaptions may be useful if coupled to sequential approaches to support dose-finding through futility interim assessments – Bayesian methods to discard minimum required activity to continue an arm are good options for implementation of such adaptions.

Multi-arm group sequential designs with a simultaneous stopping rule may be applicable in this situation, where the selection of doses was not based on a formal dose finding study (as several information was already available). This uncertainty led to a pivotal trial testing 2 different doses against placebo. A multi-arm sequential design is an alternative approach to detect as soon as possible the most efficacious dose and stop the trial (recruitment and follow up). This has many advantages, in terms of the possibility of reducing the number of patients but also to shorten

the duration of the trial, allowing open label treatment for all patients in a long term follow up separate study, without compromising the safety information and allowing faster access to patients once there's sufficient information. Since the method proposed considers a continuous outcome, an approach different to the time to event should be considered.

Given that the condition is characterised by the existence of different stages, the transition of the patient to another stage is a relevant endpoint. In PAH, this transition is defined by clinical parameters and events which may all be considered as finalist variables. Although measures of the pulmonary hypertension and functional biomarkers of the action of Opsumit are available, they are not regarded as valid endpoints as the deterioration or improvement can be measured clinically. However, there are different clinical characteristics that can define a transition from one stage to another and thus, choosing one as primary endpoint does not provide relevant data for the benefit-risk assessment. For this reason, composite endpoints that capture the global effect on the functionality of the patient are of choice in this condition.

Methods to analyse multiple endpoints like the Fallback test would be applicable if co-primary endpoints are considered, an option that could be of value when the composite endpoint is integrated by endpoints with different clinical meaning and relevance.

Bayesian approaches could be feasible also for implementing sample size reassessment or other adaptions, or for better integration of data increasing information and improving interpretation at the end of any study, especially because information on the condition is already existing and similar to the one obtained in the trial.

External controls could provide support to single arm trials, as the natural history of the disease is well known. However, the heterogeneity of patients and standards of treatment which also change over time, are a drawback to use historical controls; also, there is availability of patients to perform a placebo controlled trial, allowing proper baseline treatment.

Classic designs remain a gold standard in terms of robustness.

5 Actual development plan for Opsumit

5.1 Safety and tolerability

Opsumit (macitentan) has been studied in a total of 14 completed clinical pharmacology studies. Additional relevant PK/PD information was generated in study AC-055-201 in patients with essential hypertension and in a PK/PD sub-studies of the pivotal Phase 3 study AC-055-302 SERAPHIN and its open-label extension AC-055-303 SERAPHIN OL.

Phase I Single ascending dose 56 healthy volunteers. PK, PD, safety and tolerability. 42 macitentan treated subjects. Doses tested: 0.2, 1, 5, 25, 100, 300, 600 mg.

Phase I Multiple ascending dose. 32 healthy volunteers. PK, PD, safety and tolerability. 24 macitentan treated subjects. Once daily oral capsule 10 days: 1, 3, 10, 30.

Thorough QT study (64 healthy subjects)

Spermatogenesis (84 healthy subjects)

Mass-Balance: 6 healthy subjects

Food interaction: 10 healthy subjects

4 DDI studies: 58 healthy subjects overall

Ethnic sensitivity study: 20 healthy volunteers (10 Japanese, 10 Caucasian)

Tablet–capsule biocomparison study, 12 healthy subjects

Hepatic impairment (32 volunteers)

Renal impairment (16 volunteers)

5.2 Proof of activity/dose finding

No formal proof of concept study has been performed. Binding of an ERA to ETB receptors causes an increase in plasma ET-1 levels, which can be used as a marker of pharmacological effect and potency on the ETB receptor. This effect of ERAs is of rapid onset and thus is a PD measurement of the activity. Measures were performed in both phase I single and multiple ascending doses studies and PK-PD modelling was done.

No dedicated dose-finding study was conducted in patients with PAH. Instead, the applicant's strategy was to employ PD data on plasma ET-1 levels and hemodynamic efficacy data on blood pressure (BP) reduction in patients with mild to moderate essential hypertension to determine the doses to be tested in the Phase 3 clinical outcome study in patients with PAH. The underlying assumption was that a dose shown to be efficacious in systemic hypertension would also be hemodynamically effective in PAH, as previously observed with the ERA bosentan. Study AC-055-201 in patients with essential hypertension contributed data on the dose response for hemodynamic efficacy of macitentan.

5.3 Pivotal evidence

The authorisation is supported by a single long term pivotal phase III study (AC-055-302) which is a multicenter, randomised, double blind, placebo controlled study to assess the effects of macitentan on morbidity and mortality in patients with symptomatic pulmonary arterial hypertension.

Title of study: A Multicenter, Double-blind, Randomised, Placebo-controlled, Parallel Group, Event-driven, Phase III Study to Assess the Effects of Macitentan (ACT-064992) on Morbidity and Mortality in Patients With Symptomatic Pulmonary Arterial Hypertension.	
Investigators (Study centre): 153 study locations in 37 countries	
Studied period: 4 years First patient included on: May 2008 Last patient completed on: March 2012 (final data collection for primary endpoint) (Clintrials.gov)	**Phase of development:** Phase III
Objectives **Primary:**	

The primary objective of the study was to demonstrate that either dose (3 mg or 10 mg) of macitentan reduces the risk of morbidity and mortality in patients with symptomatic PAH. **Secondary:** • To demonstrate that either dose (3 mg or 10 mg) of macitentan improves exercise capacity, WHO functional class (FC), and reduces the risk of death due to PAH or hospitalisation for PAH up to end-of-treatment (EOT) in patients with symptomatic PAH. • To demonstrate that either dose (3 mg or 10 mg) of macitentan reduces the risk of death of all causes up to EOT and up to end-of-study (EOS). • To evaluate the safety and tolerability of macitentan in patients with symptomatic PAH.
Design: Multicenter, double-blind, randomised, placebo-controlled, event driven phase 3 trial
Number of patients, by arm: **Intended sample size:** 699. Final recruitment: 742 (250placebo/ 250 macitentan 3 mg/242 macitentan 10 mg) **Populations for analysis:** All-randomised set: 250 placebo/ 250 macitentan 3 mg/242 macitentan 10 mg

Main inclusion criteria:
• Signed informed consent prior to initiation of any study mandated procedure.
• Patients with symptomatic pulmonary arterial hypertension (PAH) in modified World Health Organization (WHO) functional class II to IV.
• Patients with the following types of pulmonary arterial hypertension (PAH) belonging to groups 1.1 to 1.3 of the Venice classification: Idiopathic (IPAH); Familial (FPAH); or Related to: Collagen vascular disease; Simple, congenital systemic-to-pulmonary shunts at least 1 year post surgical repair; Human immunodeficiency virus (HIV) infection; or Drugs and toxins.
• PAH diagnosis confirmed by hemodynamic evaluation performed prior to randomisation and showing all of the following:
  ◦ Mean pulmonary artery pressure (mPAP) > 25 mmHg at rest;
  ◦ Pulmonary capillary wedge pressure (PCWP) or left ventricular end diastolic pressure (LVEDP) < 15 mmHg; and
  ◦ Pulmonary vascular resistance (PVR) at rest >= 320 dyn×sec/cm^5.
  ◦ 6-minute walk distance (6MWD) >= 50 m.
• Men or women > 12 years of age (women of childbearing potential must have a negative pre-treatment serum pregnancy test and must use a reliable method of contraception).

Main Exclusion criteria: • PAH associated with portal hypertension, thyroid disorders, glycogen storage disease, Gaucher's disease, hereditary hemorrhagic telangiectasia, haemoglobinopathies, myeloproliferative disorders or splenectomy. • PAH associated with non corrected simple congenital systemic-to-pulmonary shunts, and combined and complex systemic-to-pulmonary shunts, corrected or non corrected. • PAH associated with significant venous or capillary involvement (PCWP > 15 mmHg), known pulmonary veno-occlusive disease, and pulmonary capillary hemangiomatosis. • Persistent pulmonary hypertension of the newborn. • Pulmonary Hypertension belonging to groups 2 to 5 of the Venice classification. • Moderate to severe obstructive lung disease: forced expiratory volume in 1 second/forced vital capacity (FEV1/FVC) < 70% and FEV1 < 65% of predicted value after bronchodilator administration. • Moderate to severe restrictive lung disease: total lung capacity (TLC) < 60% of predicted value. • Moderate to severe hepatic impairment, i.e., Child-Pugh Class B or C. • Estimated creatinine clearance < 30 mL/min

- Serum aspartate aminotransferase (AST) and/or alanine aminotransferase (ALT) > 1.5 times the upper limit of normal.
- Haemoglobin < 75% of the lower limit of the normal range.
- Systolic blood pressure < 100 mmHg.
- Acute or chronic physical impairment (other than dyspnoea), limiting the ability to comply with study requirements.
- Pregnant or breast-feeding.
- Known concomitant life-threatening disease with a life expectancy < 12 months.
- Body weight < 40 kg.
- Any condition that prevents compliance with the protocol or adherence to therapy.
- Recently started (< 8 weeks prior to randomization) or planned cardio-pulmonary rehabilitation program based on exercise.
- Treatment with endothelin receptor antagonists (ERAs) within 3 months prior to randomisation.
- Systemic treatment within 4 week prior to randomization with cyclosporine A or tacrolimus, everolimus, sirolimus (calcineurin or mammalian target of rapamycin (mTOR) inhibitors).
- Treatment with cytochrome P3A (CYP3A) inducers within 4 weeks prior to randomisation.
- Known hypersensitivity to drugs of the same class as the study drug, or any of their excipients.
- Planned treatment, or treatment, with another investigational drug within 1 month prior to randomisation.

Test product, dose and mode of administration, duration:
Macitentan 3 mg or 10 mg daily (3 and 10 mg strengths)

Reference therapy, dose and mode of administration, duration:
Active: NA
Placebo: Matching placebo
* Concomitant treatment with oral phosphodiesterase type 5 inhibitors, oral or inhaled prostanoids, calcium channel blockers, or L-arginine was allowed, provided that the patient had been receiving a stable dose for at least 3 months before randomisation. Patients receiving intravenous or subcutaneous prostanoids were excluded.

Criteria for evaluation(Definition, timing of assessments):
Main efficacy assessment:
The primary objective of reduction in the risk of a morbidity or mortality event is assessed as the time from start of treatment to the first morbidity or mortality event up to EOT, defined as follows:
Death, or onset of a treatment-emergent adverse event (AE) with a fatal outcome occurring within 4 weeks of study treatment discontinuation, or
Atrial septostomy or hospitalisation for atrial septostomy, or
Lung transplantation or hospitalisation for lung transplantation, or
Initiation of intravenous or subcutaneous prostanoids (e.g., epoprostenol, treprostinil) or hospitalisation for initiation of intravenous or subcutaneous prostanoids, or
Other worsening of PAH, defined by the combined occurrence in a patient of all the following three events:
At least 15% decrease in the 6MWD from baseline, confirmed by two 6-minute walk tests (6MWT), performed on separate days, within 2 weeks of each other.
AND
Worsening of PAH symptoms that included at least one of the following: Increase in WHO FC, or no change in patients in WHO FC IV at baseline; Appearance or worsening of signs/symptoms of right heart failure that did not respond to optimised oral diuretic therapy
AND

Need for new treatment(s) for PAH that included the following: Oral or inhaled prostanoids (e.g., iloprost); Oral phosphodiesterase inhibitors (e.g., sildenafil); Endothelin receptor antagonists (e.g., bosentan, ambrisentan) only after discontinuation of the study treatment; Intravenous diuretics.
The observation period for the primary endpoint started with first drug intake and ended at EOT + 7 days. Patients who prematurely discontinued study treatment without a morbidity or mortality event were censored at the time of study treatment discontinuation plus 7 days. Patients without an event at EOS (declared by the sponsor on 30 January 2012) were censored for the primary endpoint at their last visit in the study.
Secondary variables:
- Change in 6MWD from baseline to Month 6.
- Proportion of patients with improvement in modified WHO FC from baseline to Month 6.
- Time to death due to PAH or hospitalisation for PAH up to EOT that included: Death due to PAH (as identified by CEC) up to EOT + 7 days, or onset of a treatment-emergent AE with a fatal outcome due to PAH occurring up to 4 weeks after EOT, or Hospitalisation for PAH up to EOT + 7 days.
- Time to death of all causes up to EOT that included: Death of all causes up to EOT + 7 days, or onset of a treatment-emergent AE with a fatal outcome occurring up to 4 weeks after EOT.
- Time to death of all causes up to EOS.

Exploratory endpoints investigated the effects of macitentan on changes in 6MWD, Borg dyspnoea index and WHO FC at each assessed time-point, quality of life (QoL), and N-terminal pro-B type natriuretic peptide (NT-pro-BNP) levels.
Pharmacodynamic endpoints investigated the effects of macitentan on changes in PVR, mean right atrial pressure (mRAP), mean pulmonary pressure (mPP), cardiac index (CI), total pulmonary resistance (TPR) and mixed venous oxygen saturation from baseline to Month 6 in a sub-study to the overall protocol.
Safety:
Adverse events
Pharmacokinetic data / Pharmacodynamic data / other:
NA

Statistical methods
Analysis of efficacy:
The null hypothesis was that, independently for each dose group of macitentan (3 mg and 10 mg), there was no difference between macitentan and placebo for the risk of first occurrence of a morbidity or mortality event up to EOT (the primary endpoint). To keep the study-wise type-I error to a two-sided 0.01 'conclusive' (and highly statistically significant) level in the presence of multiple tests, each comparison of active dose versus placebo was tested at a nominal type-I error level of 0.005 (two-sided) according to Bonferroni's approach, with testing starting from the primary endpoint.
The secondary endpoints were analysed hierarchically for each dose group in the sequence of change in 6MWD (Wilcoxon rank sum test), change in WHO FC (Fisher's exact test), time to death or hospitalisation due to PAH up to EOT, and time to death of all causes up to EOT and EOS (all logrank test). No further alpha adjustment was necessary for the secondary endpoints due to the hierarchical testing procedure. No confirmatory claims can be based on variables that have a rank lower than or equal to that variable whose null hypothesis was the first that could not be rejected.
Missing data handling: No missing data imputation
Tolerability and acceptability: Description of proportions

Main results (only actual scenario)

The proportion of patients who prematurely discontinued the study was 22.4% in the macitentan 3 mg group, 16.9% in the macitentan 10 mg group, and 22.0% in the placebo group. Death was the main reason for patients not being able to complete the study in all three groups (18.8% macitentan 3 mg, 14.0% macitentan 10 mg, 17.6% placebo). Other reasons for premature discontinuation from the study included withdrawal of consent (2.4% macitentan 3 mg, 1.7% macitentan 10 mg, 1.2% placebo) and loss to follow up (1.2% macitentan 3 mg, 0.8% macitentan 10 mg, 2.8% placebo).
A confirmed primary endpoint event was recorded for 95 patients and 76 patients in the macitentan 3 mg and 10 mg groups, respectively, versus 116 patients in the placebo group (EOT + 7 days).
HR (3 mg vs placebo): 0.704 (97.5% CLs 0.516, 0.960, logrank p = 0.0108)
HR (10 mg vs placebo): 0.547 (97.5% CLs 0.392, 0.762, logrank p < 0.0001)
For the 10 mg dose it corresponded to an overall relative risk reduction of 45% and a number-needed-to-treat (NNT) of 6 patients (95% CLs 4.48, 10.80) to avoid one event at 2 years.

5.4 Supportive confirmatory efficacy and safety data

Study AC-055-303 (long term, open label extension in PAH) and AC-055B201 in patients with idiopathic pulmonary fibrosis contributed long term safety and tolerability data for the 10 mg/day dose.

5.5 Total patient exposure in the target indication

A total of 492 patients with PAH exposed to 3 and 10 mg/day.

5.6 Uncertainties/weaknesses identified

Dose finding was based of pharmacodynamic parameters on healthy volunteers and on dynamic effects in a different indication based on the development of previous drugs with the same mechanism of action. While this is useful information and allows a faster clinical development, it raises some uncertainties that are solved by testing two doses in the pivotal trial.

Main endpoint of the pivotal trial is a composite variable considering events of different clinical relevance, even if all reflect a negative clinical outcome. Individual components were analysed as secondary endpoints hierarchically and with multiplicity adjustments. When analysing mortality, no differences were observed between treatment groups.

The difference detected between the experimental treatments and placebo was already detected after 6 months of follow-up. Longer follow-up was not able to detect effects of macitentan on more finalist variables such as mortality.

Thus, it can be questioned the convenience of combining in the same composite endpoint morbidity and mortality.

Even if patients assigned to the placebo group were treated with SOC, it can be ethically questioned the value of maintaining this long follow up after the clinical benefit has been shown early after the treatment.

Finally, no stratification by WHO FC has been performed, and this would be desirable to account for clinical heterogeneity.

6 Alternative developments

6.1 Scenario 1: alternative development: Sample size reassessment and hypothesis testing in adaptive survival studies

This option was selected based on the results of the systematic exercise of applicability of methods, but also based on the clustering of conditions, reviewing the clinical characteristics of the intended therapeutic indication. For survival designs, the time to event approach was based on time to first mortality or morbidity event. As some of the clinical morbidity components of the primary endpoint can be reached early, sequential inspections are regarded as an option to optimise the duration of the trial and/or the number of patients to be included or the assignation ratio (dropping arms). Methods to optimally apply this design to obtain robust evidence have been described within the ASTERIX project.

6.1.1 Safety and tolerability

Same as main development

6.1.2 Dose finding

Same as main development

6.1.3 Pivotal evidence

Title of study: Pivotal phase III, Alternative design Option 1 A Multicenter, Double-blind, Randomised, Placebo-controlled, Parallel Group, Event-driven, Phase III Study to Assess the Effects of Macitentan (ACT-064992) on Morbidity and Mortality in Patients with Symptomatic Pulmonary Arterial Hypertension.	
Investigators (Study centre): multicentre, multinational	
Studied period: First patient included on: May 2008 Last patient completed on: will depend on the results of the group sequential analysis	**Phase of development:** Phase III
Objectives **Primary:** The primary objective of the study is to demonstrate that either dose (3 mg or 10 mg) of macitentan reduces the risk of morbidity and mortality in patients with symptomatic PAH. **Secondary:** • To demonstrate that either dose (3 mg or 10 mg) of macitentan improves exercise capacity, WHO functional class (FC), and reduces the risk of death due to PAH or hospitalisation for PAH up to end-of-treatment (EOT) in patients with symptomatic PAH. • To demonstrate that either dose (3 mg or 10 mg) of macitentan reduces the risk of death of all causes up to EOT and up to end-of-study (EOS). • To evaluate the safety and tolerability of macitentan in patients with symptomatic PAH.	
Number of patients, by arm: Patients randomised 1:1:1 stratified by centre and by WHO FC. No differences regarding the overall planned sample size with regards to the original design. A group sequential analysis for futility and overwhelming efficacy (O'Brien-Fleming type) will be done with inspections planned after 50%, 65% and 80% of events and a sample size re-assessment will be conducted at the last interim. It is expected that the interim analysis may optimise the actual number of patients by identifying at least one effective dose before the planned end of the study.	

285 events would be needed to detect a hazard ratio for macitentan/placebo of 0.55 for at least one dose group over an estimated maximum study duration of 4.1 years (using a hazard rate of 0.43 in the placebo group, an expected hazard ratio of 0.05 per year for attrition and an accrual rate of 200 patients per year). For sample size calculations, type-I error would be set to 0.005 (two-sided, Bonferroni correction to ensure an overall alpha level of 0.01) and power would be set to 90%.

Main inclusion criteria:

- Signed informed consent prior to initiation of any study mandated procedure.
- Patients with symptomatic pulmonary arterial hypertension (PAH) in modified World Health Organization (WHO) functional class II to IV.
- Patients with the following types of pulmonary arterial hypertension (PAH) belonging to groups 1.1 to 1.3 of the Venice classification: Idiopathic (IPAH); Familial (FPAH); or Related to: Collagen vascular disease; Simple, congenital systemic-to-pulmonary shunts at least 1 year post surgical repair; Human immunodeficiency virus (HIV) infection; or Drugs and toxins.
- PAH diagnosis confirmed by hemodynamic evaluation performed prior to randomisation and showing all of the following:
  - Mean pulmonary artery pressure (mPAP) > 25 mmHg at rest;
  - Pulmonary capillary wedge pressure (PCWP) or left ventricular end diastolic pressure (LVEDP) < 15 mmHg; and
  - Pulmonary vascular resistance (PVR) at rest >= 320 dyn×sec/cm^5.
  - 6-minute walk distance (6MWD) >= 50 m
- Men or women > 12 years of age (women of childbearing potential must have a negative pre-treatment serum pregnancy test and must use a reliable method of contraception).

Main Exclusion criteria:

- PAH associated with portal hypertension, thyroid disorders, glycogen storage disease, Gaucher's disease, hereditary hemorrhagic telangiectasia, haemoglobinopathies, myeloproliferative disorders or splenectomy.
- PAH associated with non corrected simple congenital systemic-to-pulmonary shunts, and combined and complex systemic-to-pulmonary shunts, corrected or non corrected.
- PAH associated with significant venous or capillary involvement (PCWP > 15 mmHg), known pulmonary veno-occlusive disease, and pulmonary capillary hemangiomatosis.
- Persistent pulmonary hypertension of the newborn.
- Pulmonary Hypertension belonging to groups 2 to 5 of the Venice classification.
- Moderate to severe obstructive lung disease: forced expiratory volume in 1 second/forced vital capacity (FEV1/FVC) < 70% and FEV1 < 65% of predicted value after bronchodilator administration.
- Moderate to severe restrictive lung disease: total lung capacity (TLC) < 60% of predicted value.
- Moderate to severe hepatic impairment, i.e., Child-Pugh Class B or C.
- Estimated creatinine clearance < 30 mL/min
- Serum aspartate aminotransferase (AST) and/or alanine aminotransferase (ALT) > 1.5 times the upper limit of normal.
- Haemoglobin < 75% of the lower limit of the normal range.
- Systolic blood pressure < 100 mmHg.
- Acute or chronic physical impairment (other than dyspnea), limiting the ability to comply with study requirements.
- Pregnant or breast-feeding.
- Known concomitant life-threatening disease with a life expectancy < 12 months.
- Body weight < 40 kg.
- Any condition that prevents compliance with the protocol or adherence to therapy.

- Recently started (< 8 weeks prior to randomization) or planned cardio-pulmonary rehabilitation program based on exercise.
- Treatment with endothelin receptor antagonists (ERAs) within 3 months prior to randomisation.
- Systemic treatment within 4 week prior to randomization with cyclosporine A or tacrolimus, everolimus, sirolimus (calcineurin or mammalian target of rapamycin (mTOR) inhibitors).
- Treatment with cytochrome P3A (CYP3A) inducers within 4 weeks prior to randomisation.
- Known hypersensitivity to drugs of the same class as the study drug, or any of their excipients.
- Planned treatment, or treatment, with another investigational drug within 1 month prior to randomisation.

Test product, dose and mode of administration, duration:
Macitentan 3 mg or 10 mg daily (3 and 10 mg strengths)

Reference therapy, dose and mode of administration, duration:
Active: NA
Placebo: Matching placebo
* Concomitant treatment with oral phosphodiesterase type 5 inhibitors, oral or inhaled prostanoids, calcium channel blockers, or L-arginine would be allowed, provided that the patient has been receiving a stable dose for at least 3 months before randomisation. Patients receiving intravenous or subcutaneous prostanoids will be excluded.

Criteria for evaluation(Definition, timing of assessments):
Main efficacy assessment:
The primary objective of reduction in the risk of a morbidity or mortality event will be assessed as the time from start of treatment to the first morbidity or mortality event up to EOT, defined as follows:
Death, or onset of a treatment-emergent adverse event (AE) with a fatal outcome occurring within 4 weeks of study treatment discontinuation, or
Atrial septostomy or hospitalisation for atrial septostomy, or
Lung transplantation or hospitalization for lung transplantation, or
Initiation of intravenous or subcutaneous prostanoids (e.g., epoprostenol, treprostinil) or hospitalization for initiation of intravenous or subcutaneous prostanoids, or
Other worsening of PAH, defined by the combined occurrence in a patient of all the following three events:
At least 15% decrease in the 6MWD from baseline, confirmed by two 6-minute walk tests (6MWT), performed on separate days, within 2 weeks of each other.
AND
Worsening of PAH symptoms that included at least one of the following: Increase in WHO FC, or no change in patients in WHO FC IV at baseline; Appearance or worsening of signs/symptoms of right heart failure that did not respond to optimized oral diuretic therapy
AND
Need for new treatment(s) for PAH that included the following: Oral or inhaled prostanoids (e.g., iloprost); Oral phosphodiesterase inhibitors (e.g., sildenafil); Endothelin receptor antagonists (e.g., bosentan, ambrisentan) only after discontinuation of the study treatment; Intravenous diuretics.
The observation period for the primary endpoint started with first drug intake and ended at EOT + 7 days. Patients who prematurely discontinued study treatment without a morbidity or mortality event were censored at the time of study treatment discontinuation plus 7 days. Patients without an event at EOS (declared by the sponsor on 30 January 2012) will be censored for the primary endpoint at their last visit in the study.
Secondary variables:
• Change in 6MWD from baseline to Month 6.
• Proportion of patients with improvement in modified WHO FC from baseline to Month 6.

• Time to death due to PAH or hospitalisation for PAH up to EOT that included: Death due to PAH (as identified by CEC) up to EOT + 7 days, or onset of a treatment-emergent AE with a fatal outcome due to PAH occurring up to 4 weeks after EOT, or Hospitalisation for PAH up to EOT + 7 days. • Time to death of all causes up to EOT that included: Death of all causes up to EOT + 7 days, or onset of a treatment-emergent AE with a fatal outcome occurring up to 4 weeks after EOT. • Time to death of all causes up to EOS. Exploratory endpoints investigated the effects of macitentan on changes in 6MWD, Borg dyspnea index and WHO FC at each assessed time-point, quality of life (QoL), and N-terminal pro-B type natriuretic peptide (NT-pro-BNP) levels. Pharmacodynamic endpoints investigated the effects of macitentan on changes in PVR, mean right atrial pressure (mRAP), mean pulmonary pressure (mPP), cardiac index (CI), total pulmonary resistance (TPR) and mixed venous oxygen saturation from baseline to Month 6 in a sub-study to the overall protocol. **Safety:** Adverse events **Pharmacokinetic data / Pharmacodynamic data / other:** NA
Statistical methods **Primary efficacy analysis:** A group sequential analysis will be done with inspections planned after 50% and 75% of events and a sample size re-assessment will be conducted at the last interim. A Mantel-Haenzel test will be done for comparing the proportion of patients free of morbi-mortality events at 6 months follow-up between groups, adjusted by stratification factors. O'Brian-Fleming adjustment for multiplicity will be applied to account for multiple inspections. Stopping rules for both futility and superiority will be determined by the result of the O'Brian-Fleming adjustment to account for a maximum level of an overall one-sided type 1 error of 0.5% (overall 1%, 0.5% Bonferroni-adjusted per each dose vs placebo, two-sided). **Secondary efficacy end-points** at different times will be analysed similarly, and survival methods will be applied to the analysis of time to event (log-rank and Kaplan-Meier). **Missing data handling:** No missing data will be imputed. A sensitivity analysis by imputing failure at the time of censoring for treatment related missingness will be done. **Tolerability and acceptability:** Descriptive analysis of the proportions of TEAE, TEAEs that led to death (component of the primary endpoint).

6.1.4 Supportive confirmatory efficacy data

Same as main development

6.1.5 Supportive safety data

Same as main development

6.2 Scenario 2: alternative development: use of co-primary endpoints

This option was selected based on the results of the systematic exercise of applicability of methods, but also based on the clustering of conditions, reviewing the clinical characteristics of the intended therapeutic indication.

Main endpoint of the pivotal trial is a composite variable considering events of different clinical relevance even if all reflect a negative clinical outcome. Individual components would be analysed as secondary endpoints hierarchically and with multiplicity adjustments. A

different approach based on co-primary endpoints can be explored and methods for the analysis of multiple endpoints as described in the ASTERIX project can be implemented.

6.2.1 Safety and tolerability

Same as main development

6.2.2 Dose finding

Same as main development

6.2.3 Pivotal evidence

Title of study: Pivotal phase III, Alternative design Option 2 A Multicenter, Double-blind, Randomised, Placebo-controlled, Parallel Group, Event-driven, Phase III Study to Assess the Effects of Macitentan (ACT-064992) on Morbidity and Mortality in Patients with Symptomatic Pulmonary Arterial Hypertension.	
Investigators (Study centre): multicentre, multinational	
Studied period: First patient included on: May 2008 Last patient completed on: Same as actual development.	**Phase of development:** Phase III
Objectives **Primary:** The primary objective of the study is to demonstrate that either dose (3 mg or 10 mg) of macitentan reduces the risk of morbidity in patients with symptomatic PAH. **Secondary:** • To demonstrate that either dose (3 mg or 10 mg) of macitentan improves exercise capacity, WHO functional class (FC), and reduces the risk of death due to PAH or hospitalisation for PAH up to end-of-treatment (EOT) in patients with symptomatic PAH. • To demonstrate that either dose (3 mg or 10 mg) of macitentan reduces the risk of death of all causes up to EOT and up to end-of-study (EOS). • To evaluate the safety and tolerability of macitentan in patients with symptomatic PAH.	
Number of patients, by arm: Patients randomised 1:1:1 stratified by centre and by WHO FC The individual sample size for each of the 3 (co-)primary end points, (i) ≥15% decrease in the 6MWD from baseline, (ii) increase in WHO FC, (iii) signs or symptoms of right-sided heart failure, is 306, 220 and 215 per arm. The expected placebo vs best active dose rates of events are (i) 48.4% vs 33.9%, (ii) 44.8% vs 28.1% and (iii) 50.4% vs 33.1%, respectively. The sample sizes take into account a Bonferroni adjustment to split the overall 1% two-sided alpha for each active dose versus placebo and a 90% statistical power. About 220 per arm is the recommended sample size for this design. This leads to a similar sample size to the one of the original planned but increases the chances of finding at least one significant end-point by the use of Fallback-tests.	
Main inclusion criteria: • Signed informed consent prior to initiation of any study mandated procedure. • Patients with symptomatic pulmonary arterial hypertension (PAH) in modified World Health Organization (WHO) functional class II to IV. • Patients with the following types of pulmonary arterial hypertension (PAH) belonging to groups 1.1 to 1.3 of the Venice classification: Idiopathic (IPAH); Familial (FPAH); or Related to: Collagen vascular disease; Simple, congenital systemic-to-pulmonary shunts at least 1 year post surgical repair; Human immunodeficiency virus (HIV) infection; or Drugs and toxins.	

- PAH diagnosis confirmed by hemodynamic evaluation performed prior to randomiSation and showing all of the following:
  - Mean pulmonary artery pressure (mPAP) > 25 mmHg at rest;
  - Pulmonary capillary wedge pressure (PCWP) or left ventricular end diastolic pressure (LVEDP) < 15 mmHg; and
  - Pulmonary vascular resistance (PVR) at rest >= 320 dyn×sec/cm^5.
  - 6-minute walk distance (6MWD) >= 50 m.
- Men or women > 12 years of age (women of childbearing potential must have a negative pre-treatment serum pregnancy test and must use a reliable method of contraception).

Main Exclusion criteria:

- PAH associated with portal hypertension, thyroid disorders, glycogen storage disease, Gaucher's disease, hereditary hemorrhagic telangiectasia, hemoglobinopathies, myeloproliferative disorders or splenectomy.
- PAH associated with non corrected simple congenital systemic-to-pulmonary shunts, and combined and complex systemic-to-pulmonary shunts, corrected or non corrected.
- PAH associated with significant venous or capillary involvement (PCWP > 15 mmHg), known pulmonary veno-occlusive disease, and pulmonary capillary hemangiomatosis.
- Persistent pulmonary hypertension of the newborn.
- Pulmonary Hypertension belonging to groups 2 to 5 of the Venice classification.
- Moderate to severe obstructive lung disease: forced expiratory volume in 1 second/forced vital capacity (FEV1/FVC) < 70% and FEV1 < 65% of predicted value after bronchodilator administration.
- Moderate to severe restrictive lung disease: total lung capacity (TLC) < 60% of predicted value.
- Moderate to severe hepatic impairment, i.e., Child-Pugh Class B or C.
- Estimated creatinine clearance < 30 mL/min
- Serum aspartate aminotransferase (AST) and/or alanine aminotransferase (ALT) > 1.5 times the upper limit of normal.
- Hemoglobin < 75% of the lower limit of the normal range.
- Systolic blood pressure < 100 mmHg.
- Acute or chronic physical impairment (other than dyspnea), limiting the ability to comply with study requirements.
- Pregnant or breast-feeding.
- Known concomitant life-threatening disease with a life expectancy < 12 months.
- Body weight < 40 kg.
- Any condition that prevents compliance with the protocol or adherence to therapy.
- Recently started (< 8 weeks prior to randomization) or planned cardio-pulmonary rehabilitation program based on exercise.
- Treatment with endothelin receptor antagonists (ERAs) within 3 months prior to randomisation.
- Systemic treatment within 4 week prior to randomization with cyclosporine A or tacrolimus, everolimus, sirolimus (calcineurin or mammalian target of rapamycin (mTOR) inhibitors).
- Treatment with cytochrome P3A (CYP3A) inducers within 4 weeks prior to randomisation.
- Known hypersensitivity to drugs of the same class as the study drug, or any of their excipients.
- Planned treatment, or treatment, with another investigational drug within 1 month prior to randomisation.

Test product, dose and mode of administration, duration:
Macitentan 3 mg or 10 mg daily (3 and 10 mg strengths)

Reference therapy, dose and mode of administration, duration:
Active: NA

Placebo: Matching placebo
* Concomitant treatment with oral phosphodiesterase type 5 inhibitors, oral or inhaled prostanoids, calcium channel blockers, or L-arginine would be allowed, provided that the patient has been receiving a stable dose for at least 3 months before randomisation. Patients receiving intravenous or subcutaneous prostanoids would be excluded.

Criteria for evaluation(Definition, timing of assessments):
Main efficacy assessment:
The following primary 3 (co-)primary endpoints related to the worsening of PAH are considered as events:

- At least 15% decrease in the 6MWD from baseline, confirmed by two 6-minute walk tests (6MWT), performed on separate days, within 2 weeks of each other.
- Increase in WHO FC.
- Signs or symptoms of right-sided heart failure.

Secondary variables:

- Change in 6MWD from baseline to Month 6.
- Proportion of patients with improvement in modified WHO FC from baseline to Month 6.
- Time to death due to PAH or hospitalisation for PAH up to EOT that included: Death due to PAH (as identified by CEC) up to EOT + 7 days, or onset of a treatment-emergent AE with a fatal outcome due to PAH occurring up to 4 weeks after EOT, or Hospitalisation for PAH up to EOT + 7 days.
- Time to death of all causes up to EOT that included: Death of all causes up to EOT + 7 days, or onset of a treatment-emergent AE with a fatal outcome occurring up to 4 weeks after EOT.
- Time to death of all causes up to EOS.

Exploratory endpoints investigated the effects of macitentan on changes in 6MWD, Borg dyspnea index and WHO FC at each assessed time-point, quality of life (QoL), and N-terminal pro-B type natriuretic peptide (NT-pro-BNP) levels.
Pharmacodynamic endpoints investigated the effects of macitentan on changes in PVR, mean right atrial pressure (mRAP), mean pulmonary pressure (mPP), cardiac index (CI), total pulmonary resistance (TPR) and mixed venous oxygen saturation from baseline to Month 6 in a sub-study to the overall protocol.
Safety:
Adverse events
Pharmacokinetic data / Pharmacodynamic data / other:
NA

Statistical methods
Primary efficacy analysis: The primary efficacy analysis will consider the 3(co-)primary endpoints, and they will be analysed with Fallback tests for co-primary endpoints. A Bonferroni adjustment will be used to split the overall 1% two-sided alpha for each active dose versus placebo.
Secondary efficacy end-points at different times will be analysed similarly, and survival methods will be applied to the analysis of time to event (log-rank and Kaplan-Meier).
Missing data handling: No missing data will be imputed. A sensitivity analysis by imputing failure at the time of censoring for treatment related missingness will be done.
Tolerability and acceptability: Descriptive analysis of the proportions of TEAE, TEAEs that led to death (component of the primary endpoint) would be performed.

6.2.4 Supportive confirmatory efficacy data

Same as main development

6.2.5 Supportive safety data

Same as main development

7 Analysis of the practical, ethical and regulatory impact

7.1 Analysis of option 1

Method assessed: Option 1: Sample size reassessment and hypothesis testing in adaptive survival studies	Improves?	Comments:
Practical considerations		
• **May reduce sample size requirements**	Yes	It is considered that, depending on the magnitude of the effect, group-sequential designs can achieve about 30% reduction of sample size, but may also increase study size in some circumstances
• **May shorten time to study completion**	Yes	Theoretically yes, it may depend on the sample size reassessment (if able to downsize the number of patients) or the shortening of the treatment and follow-up treatment by prematurely stopping the double blind phase once the objective of the trial has been reached
• **May ease recruitment**	Depends	Patients will be assigned only 1/3 to placebo and also the reduction in sample size migh ease recruitment; also, the possibility of limiting the time exposed to placebo and
Statistical assessment:		
• **Improves internal validity**	Partially	The Pivotal trial had already a fair internal validity
• **Increases stability of estimates**	No	
• **Increases sensibility to changes**	Partially	Although same variable, placebo control is more stable (less changes in 6 – 12 months) and theoretically enhances sensitivity to treatment-related changes
• **Compliant with predetermination**	Yes	Conditions for Interim Analysis are pre-established
• **Consistency (discuss)**	No (may worsen)	It all depends on the ability to show an effect in secondary endpoints if a reduced sample size or shorter term Follow-Up; this is less a problem if follow-up is extended after completion of main end-point
• **Robustness of method (discuss)**	No	The pivotal trial is already a robust method
• **Protection against type I and II errors (discuss)**	No	Maintained but not improved. Multiplicity adjustments for group sequential designs for small populations, and Randomized Clinical Trial methodology
Regulatory assessment:		

• **Risk of bias and credibility**	Yes	By stopping the trial early and providing less long term follow-up in a double blind fashion, long term endpoints such as mortality cannot be assessed properly
• **External validity (discuss)**	Similar	Similar than previously
• **Therapeutic positioning and comparisons**	Yes	The approach is an add-on to other treatments for the condition targeting different mechanisms; even if there are other treatments with the same mechanism of action available, no comparison is done and thus no therapeutic positioning can be concluded
• **Informative on relevance and clinical impact**	Yes	Similar or less than previously as no information on the compound of the composite variable that take long term to be modified could be missed
• **Enough information on safety**	Yes	Although shorter safety data from double blind, long term follow-up is maintained in a similar or larger number of patients
• **Suitable information for risk-benefit balance**	Yes	
Ethical assessment:		
• **May minimise risks**	Yes	By reducing sample size and duration of ineffective treatments
• **May maximise access to treatment**	Yes	Yes, by accessing earlier (in case that effectiveness is shown) to the long term open label extension
• **May minimise unnecessary exposure to ineffective treatments or placebo**	Yes	By discontinuing experimental treatment early if it does not show efficacy and stopping placebo arm once the efficacy has been shown
• **Considers patient input**	No	Endpoints in this condition are widely recognized

7.2 Analysis of option 2

Method assessed: Option 2: Use of co-primary endpoints	Improves?	Comments
Practical considerations		
• **May reduce sample size requirements**	No	
• **May shorten time to study completion**	No	Depending on the variables chosen as co-primary endpoints
• **May ease recruitment**	No	
Statistical assessment:		

• **Improves internal validity**	No change	
• **Increases stability of estimates**	No	
• **Increases sensibility to changes**	No	
• **Compliant with predetermination**	Yes	
• **Consistency (discuss)**	No (may worsen)	
• **Robustness of method (discuss)**	Yes	
• **Protection against type I and II errors (discuss)**	Yes	Statistical analysis will consider the use of Fallback tests for co-primary endpoints. Additionally, Bonferroni adjustment will be used to split the overall 1% two-sided alpha for each active dose versus placebo. Sample size was estimated for a type II error of 90%.
Regulatory assessment:		
• **Risk of bias and credibility**	Yes	Even if the method allows for inferences in settings where only some of the co-primary endpoints show a significant effect, regulators may be reluctant to accept this innovative methodology
• **External validity (discuss)**	No	Similar than previously
• **Therapeutic positioning and comparisons**	No	This approach may help to better define the effects of the treatment on different endpoints in a robust fashion, but no for comparison with other active treatments
• **Informative on relevance and clinical impact**	Yes	The impact on each of the clinical endpoints is assessed separately in a non hierarchical way
• **Enough information on safety**	No	The same as previous studies
• **Suitable information for risk-benefit balance**	Yes	This design will provide results of the clinical co-primary endpoints globally and separately.
Ethical assessment:		
• **May minimise risks**	No	No changes foreseen

• **May maximize access to treatment**	No	No changes foreseen
• **May minimise unnecessary exposure to ineffective treatments or placebo**	No	No changes foreseen
• **Considers patient input**	Yes	Patient reported outcomes or GAS may be considered as co-primary endpoints

8 Recommendations

The development of Opsumit for the treatment of PAH is characteristic of a non-oncological disease fitting the ASTERIX cluster 6 (staged conditions). There are a number of treatments available for the treatment of this condition and a regulatory guideline for the development of new products was issued by EMA, acknowledging that it is a disease with very active research and this may not be the standard situation in orphan conditions. Analysing in depth the development actually performed and possible alternatives may establish a set of recommendations for diseases with similar characteristics also classified within this ASTERIX cluster to guide developments.

We have maintained the overall 1% two-sided type I error due to the fact of only one-single pivotal study as in the original development plan, however, this might be possible be relaxed a bit up to the standard 5%. A prospective defined meta-analysis with pre- plus post-MAA studies would have been probably helpful to alleviate the concerns of a single-pivotal study and thus the sample size might be optimised.

The use of sequential methods and adaptive approaches are recommended in general in this cluster. The use of time to event endpoints is also a characteristic of conditions classified as staged diseases, thus, the proposal of applying methods for sample size reassessment and hypothesis testing in adaptive survival studies is a recommendation that could be valid in most situations.

Furthermore, stratification by the different stages of the condition is necessary in case where patients in different stages are included in the studies.

Also, the use of more than one endpoint is an option to obtain a robust evidence of the effect of the treatment.