Economic relevance and validation procedure

Orphan Medicinal Products

- *Determination of relevant issues in the cost-effectiveness evaluation*

- *Role of cost-effectiveness evaluation and other economic instruments in current coverage decisions*

- *Guidance on coverage decision-making in orphan disease: an algorithmic approach*

Economic relevance and validation procedure

Orphan Medicinal Products

- *Determination of relevant issues in the cost-effectiveness evaluation*

- *Role of cost-effectiveness evaluation and other economic instruments in current coverage decisions*

- *Guidance on coverage decision-making in orphan disease: an algorithmic approach*

Colophon

Expert committee
<u>Coordinator expert committee:</u>
M. (Mark) J.C. Nuijten, MD, PhD, MBA   Amsterdam,
                                                    The Netherlands

<u>Expert committee members:</u>
Prof. J (John) Hutton, BPhil, Health Economist   University of York, York,
                                                    United Kingdom

Prof. F. (Frans) H. Rutten, PhD, Health Economist   Institute for Medical Technology Assessment
                                                    (iMTA), EUR Rotterdam,
                                                    The Netherlands

P. (Peter) M. van Hasselt, MD, PhD,   University Medical Centre Utrecht,
Pediatrician for metabolic diseases   for metabolic diseases,
                                                    The Netherlands

C. (Cor) Oosterwijk, PhD   Dutch Genetic Alliance (VSOP), Soest,
                                                    The Netherlands

D. (Dominique) J. Dubois, MD, FFPM, FBCPM   Université Libre de Bruxelles (ULB),
Physician Specialist in Pharmaceutical Medicine   Belgium

*The experts have contributed equally to this report. A declaration of interest by the aforementioned
experts has been included in this report.*

Advisors
Mrs. H. (Hanka) K. Meutgeert   VKS (Patient – Association for metabolic
                                                    disorders), Zwolle,
                                                    The Netherlands

Europe-Expro

Europe-ExPro identifies medical and pharmaceutical issues requiring assessment of their relevance by experts. Assessment of relevance can be done from a clinical, economic and/or patient perspective. For this purpose, Europe-Expro facilitates expert procedures. The outcomes of expert procedures are published as a report for external use and are summarised in a manuscript that will be submitted for publication.

An expert procedure is performed by an expert committee, without any influence and input of external stakeholders, such as medical and pharmaceutical companies, and policy institutes or payers.

This report is initiated on the observation that cost-effectiveness evaluation of orphan medicinal products is confronted with large confidence intervals in incremental cost-effectiveness ratios (ICERs) or extremely high ICERs and therefore rejection for uptake in the health insurance package (coverage) by health authorities in Europe. This observation required more insight into what drives these high ICERs and what policies may support the appropriate use of orphan medical products.

For more information on this expert procedure, please contact: Europe-Expro
Zeppelinstraβe 73, D-81669 Munich
Tel.: +49 1717577031; E.: info@europe-expro.eu; W.: www.europe-expro.eu

EUROPE-EXPRO

EXPERT PROCEDURES ON CLINICAL, ECONOMIC, AND PATIENT RELEVANCE

Organisations involved

The expert procedure is supported by the Dutch Genetic Alliance (VSOP, Association for cooperating parent and patient organisations in the field of rare and genetic diseases, the Netherlands) and the patient organisation VKS (Adults, Children and Metabolic disorders; the Netherlands).

Funds

Europe-ExPro has raised a fund, called 'R&D, Relevance Development'. Europe-ExPro receives unrestricted grants and funds of companies and institutes that support the development of relevance procedures. Europe-ExPro maintains a zero-policy, which means that fund and grant suppliers do not have any interaction, input or influence on the expert procedures, nor are grant and fund suppliers involved in reviewing the outcomes of the expert procedures. Grants and funds are used for researching disease areas and issues that need expert views on relevance, for executing the expert procedures, coordinating the expert procedures, writing reports and publishing them. The zero-policy applies also to other organisations or institutes, like health insurers, professional societies, and health authorities, which have an interest in a particular expert procedure.

Information on this expert procedure is available on request. For this expert procedure unrestricted grants have been received from Genzyme Europe B.V. and Novartis Pharma B.V..

Summary

Introduction
This report reflects the outcomes of an expert procedure, which was initiated on the observation that cost-effectiveness evaluation of orphan medicinal products is confronted with a large confidence interval on the incremental cost-effectiveness ratios (ICERs), or extremely high ICERs and therefore rejection of products for uptake in the health insurance package (coverage) by health authorities in Europe. This observation required more insight into what drives these high ICERs and what policies may support the appropriate use of orphan medicinal products. Europe-Expro facilitated this expert procedure.

Objective
Based on the above mentioned observation, an expert committee addressed the following questions:

> *"Which clinical and economic issues complicate cost-effectiveness evaluation of orphan diseases and orphan medicinal products?"*
>
> *"What is the role of cost-effectiveness in coverage decisions for orphan medicinal products and how to deal with the exceptional outcomes of cost-effectiveness evaluation?"*
>
> *"Should arguments (criteria) other than cost-effectiveness be given the main emphasis in coverage assessment procedures of orphan medicinal products, when the clinical data (effectiveness) of a medicinal product at time of launch show limitations".*

Although we are aware of the fact that limitations of the clinical data at time of launch are not exclusively the case for orphan medicinal products, we have selected them as a group to address the above mentioned questions.

Background
Traditional coverage systems and assessment of the value for money of new medicinal products, is based on standard coverage methods for traditional medicinal products, which entered the market in the seventies and eighties (e.g. antibiotics, antidepressants) and which had a relative limited daily price. However, since the late nineties, new classes of medicinal products have entered the market, which offer treatment for a small number of patients at a high price per patient. Among these products are, for example, biologicals with a completely different mechanism of action, new oncology therapies and orphan medicinal products, all (extremely) high priced medicinal products with a high ICER or large confidence interval on the ICER caused by limitations of the clinical data at the time of launch. As a result, some of these products were rejected for coverage, while others were accepted for coverage, often in decision processes lacking transparency and a clear formalised structure or framework, which justifies the criteria, arguments and final decision. In response to these issues, stakeholders have disputed the appropriateness of the main emphasis that is put on the

cost-effectiveness criterion in national coverage assessment procedures. They initiated the development of alternative assessment frameworks (MCDA), with a lower weighting of the incremental cost-effectiveness outcome in the overall coverage decision and the addition of other criteria, measuring for example social values, and alternative policy tools (e.g. prescription guidelines, stratified medicine, value based pricing, risk sharing and outcomes research and temporary coverage). These alternative coverage instruments might seem relevant and useful for the coverage of expensive medications, including orphan medicinal products, but are still in an evaluation process with no final judgment being reached as to their appropriateness.

Methods
We listed the most relevant clinical and economic issues that are perceived to complicate the cost-effectiveness evaluation of orphan diseases and orphan medicinal products: sample size, stratification, outcome measures, patient benefits, direct versus indirect costs. Most of the above mentioned issues originate from the "small numbers" of patients available for study, which is the main factor complicating cost-effectiveness evaluation. Therefore, medicinal products, whether they are orphan or non-orphan, that are hampered by limitations to the clinical data at the time of launch, may require a different approach in coverage decisions and HTA assessment procedures.
We discussed two possible solutions: 1) circumvention of the standard assessment criterion -cost-effectiveness- by adding other criteria, measuring for example social values, or 2) maintenance of the standard assessment criterion -cost-effectiveness- and putting effort into optimisation of the utilisation of products, for example limiting their use to those patients that show maximum benefit and in such a way (containing stringent starting and stopping rules) that the associated cost-effectiveness is still within reasonable limits or at least improves substantially.

Outcomes
Although alternative assessment frameworks and policy tools might be designed well, and although such a systematic and transparent system might help frame a more structured dialogue between stakeholders and health authorities and payers, we questioned the potential of the MCDA approach and policy tools to provide helpful answers and to be sustainable, given the importance of cost containment in health care. Ultimately, these alternative assessment frameworks and policy tools do not solve the basic problems of 'lack of strong or convincing clinical evidence', and 'high price of the medicinal product', both resulting in a large confidence interval in the ICER and a high ICER.
Therefore, instead of creating new instruments, we suggest that efforts should be directed at optimising the input for the cost-effectiveness evaluation, for example by limiting the use of these medicinal products to those patients that show maximum benefit and in such way (containing stringent start and stopping rules) that the associated cost-effectiveness is still within reasonable limits or at least improves substantially.
In this report we provide guidance on coverage decisions for (orphan) medicinal products that show limitations of the clinical data at the time of launch, by developing a Simple Transparent Algorithmic Multidisciplinary Procedure (STAMP) for the purpose of optimising the input for the standard set of decision criteria (effectiveness, cost-effectiveness and budget impact).

Abbreviations

BIA	Budget impact analysis
CHMP	Committee for Medicinal Products for Human Use
CVZ	National Health Insurance Board (Dutch: College voor zorgverzekeringen)
EMA	European Medicines Agency
EU	European Union
EUCERD	European Union Committee of Experts on Rare Diseases
EUnetHTA	European Network of Health Technology Assessment
HRQoL	Health Related Quality of Life
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
MCDA	multi-criteria decision analysis
MPS I	Mucopolysaccharidose I
MTHFR deficiency	Methylenetetrahydrofolate reductase deficiency
NHS	National Health Service
NICE	National Institute of Health and Care Excellence
RCT	Randomised controlled trials
SM	Stratified Medicine
SMC	Scottish Medicines Consortium
STAMP	Simple Transparent Algorithmic Multidisciplinary Procedure
UK	United Kingdom
VBP	Value Based Pricing
QALY	Quality-Adjusted Life Year
QoL	Quality of Life

Table of content

1. Determination of relevant issues in the cost-effectiveness evaluation of orphan medicinal products

1.1 Introduction

The decision of health authorities on coverage of a medicinal product in the health insurance package is based on the value for money of that medicinal product. Health authorities will make a trade-off between the incremental clinical benefit and the extra cost of the new medicinal product versus standard therapy. Until recently, the judgment of the clinical benefit of medicinal products was based on traditional clinical trial outcomes criteria (efficacy, safety and quality) used for marketing-authorisation (registration) and a subjective value judgment based on relative prices. Today, more formal methods of health technology assessment (HTA), such as budget impact and cost-effectiveness analysis and more standardised measures of patient benefit are applied in order to make a value for money decision. This has led to extra clinical data requirements which all relate to the use of the medicinal product in real daily practice. This contrasts with the traditional clinical trial outcomes derived from randomised controlled trials (RCT's) in well controlled settings. The current most important new data requirements for coverage decisions by health authorities are:

- Effectiveness which is the actual health benefit achieved by a medicinal product in daily practice. This contrasts with efficacy, which is the benefit measured under ideal conditions in a homogeneous group of patients.
- Cost-effectiveness of a medicinal product is based on the costs that will result from its use, and the potential savings that will be made compared with other products and/or treatments and the health benefits to the patient. The most used measure of this health benefit is the "quality-adjusted life year," or QALY. The lower the ratio of cost per QALY gained, the more cost-effective a health intervention is said to be. In many countries threshold values are used in coverage decisions. Medicinal products with ratios above these threshold values are not given coverage under the national health insurance package. These threshold values vary considerably from country to country. Values ranging from $50,000 to $100,000 per QALY gained (€37,000 to €74,000) are sometimes used as a threshold in the United States (US)[1], whereas in the United Kingdom (UK) NICE (National Institute of Health and Care Excellence) has adopted a cost effectiveness threshold range of £20,000 to £30,000 per QALY gained (€23,800 to €35,700). In the case of end-of-life treatment an ICER up to £55,000 (€65,450) may be accepted.[2] Other proposals include a differential threshold value between diverse disease and treatment characteristics, for example in The Netherlands, recently a range between €20,000 and €80,000 per QALY[3] gained has been suggested depending on burden of disease for the patient (high burden, high cost per QALY threshold). There is currently an increasing demand for health economic data in the decision-making process in Europe. Several countries now have formal requirements (e.g. England/Wales, Scotland, Sweden, The Netherlands, Belgium, Finland and Portugal).

- Budgetary impact data from a financial analysis illustrate the impact of a new medicinal product on the annual national medicinal product budget and total health care budget.

Cost-effectiveness data should permit reliable, reproducible and verifiable insight into the effectiveness of a medicinal product, the costs that will result from its use, and the possible savings that will be made compared with other medicinal products and/or treatments. For several years, stakeholders have disputed the appropriateness of the cost-effectiveness criterion in national coverage assessment procedures in cases where the clinical data at time of launch show limitations. The associated uncertainty may result in a large confidence interval, which may induce authorities to make a negative coverage decision. Consequently, the risk of exclusion of these medicinal products from coverage is high. In some cases, it is already clear upfront that the incremental cost-effectiveness ratio (ICER) will never fit into any threshold range, even if the clinical data would become more convincing. If at the same time these products are the only available treatment for that specific disease, the question arises whether other criteria than cost-effectiveness should be given the main emphasis in coverage assessment procedures for such medicinal products.

Authorities like EMA and NICE have demonstrated that assessment of orphan medicinal products with limitations of the clinical data at time of launch is feasible when standard assessment criteria are applied. Although the EMA states in the Orphan Regulation that no special assessment for marketing authorisation of orphan medicinal products exists, and that such products should be subject to the same assessment as other medicinal products, to guarantee that there will be no 'second rate' assessment, they have a 'Guideline on Clinical Trials in small populations'[4]. In this way, the assessment criteria remain standardised but flexibility is provided in the interpretation of the benefits of medicinal products with limitations in the availability of clinical data. Based on this approach the Committee for Medicinal Products for Human Use (CHMP) has been able to grant Marketing-authorisation for both orphan and non-orphan medicinal products with limitations in the clinical data at time of launch. In such cases, the CHMP requests additional information to be submitted within an agreed period of time. This additional information is submitted for re-evaluation of the marketing authorisation status of the particular orphan or non-orphan medicinal product. In a similar way NICE sticks to its standard assessment criteria, but at the same time allows the use of best available evidence, not always RCT's. When a product is approved with limited clinical evidence NICE is likely to set an early review date for the appraisal, at which point more robust data would be expected.

Questions have been raised by stakeholders, such as manufacturers and patient organisations, about the appropriateness of the cost-effectiveness criterion in national coverage assessment procedures in cases where the clinical data at time of launch show limitations. These disputes have been exacerbated by the lack of transparency, in some countries, in the relationship between cost-effectiveness and health authorities' decisions for covering medicinal products, and the lack of a clear formalised structure or framework, which justifies the criteria, arguments and final coverage decision by health authorities. Consequently the following questions are addressed in this report:

"Which clinical and economic issues complicate cost-effectiveness evaluation of orphan diseases and orphan medicinal products?"

"What is the role of cost-effectiveness in coverage decisions for orphan medicinal products and how to deal with the exceptional outcomes of cost-effectiveness evaluation?"

"Should arguments (criteria) other than cost-effectiveness be given the main emphasis in coverage assessment procedures of orphan medicinal products, when the clinical data (effectiveness) of a medicinal product at time of launch show limitations".

Although we are aware of the fact that limitations of the clinical data at time of launch are not exclusively the case for orphan medicinal products, we have selected orphan medicinal products as a group in which to address the above mentioned questions.

In order to address the above questions, we focus in this chapter on the main clinical and also economic issues that currently cause dispute in the cost-effectiveness evaluation of orphan medicinal products. The issues are discussed in section 1.3 and 1.4 of this chapter. Prior to this discussion we elaborate on the definition and characteristics of orphan medicinal products in general. The second and third questions are addressed in chapters 2 and 3 respectively.

1.2 Orphan disease

Orphan disease is defined in the EU Orphan Regulation 141/2000[5] as:
- It must be a disease that is life-threatening or chronically debilitating;
- The prevalence of the condition in the EU must not be more than 5 in 10,000 or it must be unlikely that the marketing of a medicine for its treatment would generate sufficient returns to justify the investment needed for its development;
- No satisfactory method of diagnosis, prevention or treatment of the condition concerned has been authorised in the EU.

Additionally, other factors like heterogeneity of patient population, multiple affected organ systems, and broader health and financial impacts on the family and carers of the patient complicate many orphan diseases. Some orphan diseases have the problem that variety is present on many levels within the disease itself, namely in health state, in severity of the disease, and in presented symptoms.

It is important to consider that each separate issue in itself or the combination of issues might not be unique for orphan diseases. We are aware of non-orphan diseases, for example some cancers or multiple sclerosis, that also might be affected by the same (combination of) issues. Thus, the

treatment of these types of diseases may require a different approach in coverage decisions and HTA assessment procedures. However, the key problem of small numbers of patients for study (and to generate future revenue) makes the difficulties more acute in orphan diseases

1.3   Relevant *clinical* issues for orphan medicinal products and diseases

In this section, we describe the most relevant clinical issues for consideration in the application of cost-effectiveness evaluation of orphan medicinal products as part of the decision-making process for coverage under the national health insurance package. Although, as stated above, almost all these clinical issues are not unique to orphan diseases and orphan medicinal products, the problems are concentrated in these diseases and products, which therefore provide a suitable framework within which to discuss their resolution.

Sample size and heterogeneity
The sample size of clinical trials and observational studies is limited in orphan disease, because of 1) a limited number of potential patients with a specific disease; 2) the spectrum of disease severity within an individual disease may be enormous, ranging from early infancy to adulthood and from life threatening to mildly disabling; 3) patients to be included are in various stages of the disease at the time treatment can be initiated; and 4) there is a lack of historical epidemiological data which would allow an estimation of the expected course of the disease (without treatment), and a lack of biomarkers to monitor disease progression.
As a consequence of this, it is often not possible to recruit a sufficiently large, homogenous group of patients, such as only pre-symptomatic patients, or only moderate-severe patients, to test hypotheses using conventional statistical methods. Thus within study populations patients may vary considerably in the severity of the disease, duration of the disease, as well as in other prognostic factors such as age, disease progression and type of symptoms, as most orphan diseases have an impact on various organ systems.

Example[6]
In Pompe disease, patients may differ substantially in the severity of the disease and the type of symptoms.

Paediatric population:
• A three-month old baby is hospitalised with breathing problems due to a heavy cold. Radiographs of the chest of the child show that the heart is greatly enlarged.
• A two-year old child, who has just learned to walk, staggers and appears to walk with a limp. As a baby the child had difficulty rolling over and lifting the head.

Adult population:
• A 20-year-old woman has shortness of breath and gets pain in her muscles as she runs upstairs.
• A 37-year-old man falls asleep during the day and has trouble breathing when he lies down.

The same is the case in other lysosomal storage disorders, such as MPS I. These examples illustrate the need for stratification of the patient population, even within similar age groups. This heterogeneity, leads to extra variance in outcomes and consequently difficulties in demonstrating the statistical significance of the clinical outcomes.

Stratification
There is a constant trade-off between overall sample size and the heterogeneity of the patient population. Restricted patient numbers mean that the traditional approaches to stratification are unlikely to produce meaningful results. More research is required in diagnostic testing, biomarkers and clinical efficacy and quality of life 'markers', corresponding with the principles of stratified medicine in order to identify the at risk persons for developing the disease or defining the persons that benefit most from the therapy.

Example[6]
Stratification into subpopulations within the non-classical patients with Pompe disease may be questioned, especially because there are power problems due to low number of patients. An important issue is the type of patients in the non-classical study population of Pompe and the predictive value of diagnostic testing. A positive test may not predict that a patient will ever really develop Pompe disease. Therefore treatment based on a positive test may be redundant in a proportion of patients and may lead to an unnecessarily high impact on the medicinal product budget. However, if you do not treat the "patient", and he/she develops Pompe disease, the opportunity of treatment in an earlier stage may be lost. The same issue of predictive value of diagnostic testing applies to Fabry disease. In these cases, the decision for "treatment" or "no treatment" may therefore also lead to the dilemma that spending money on orphan treatments with no proven benefit is denying benefits to patients with other conditions and evidence of benefit.

Study design
Although the RCT (randomised controlled trial) is likely to produce the least biased clinical efficacy data, a RCT in orphan diseases cannot always be performed due to, for example, sample size constraints or ethical reasons against a 'best supportive care'. Randomised uncontrolled clinical studies, open label studies and also registries, serving as best available evidence may be alternative options. In these non-RCT's the sample size issue may be partially solved, because some heterogeneity can be controlled for in statistical analysis. However, the use of non-RCT's, especially registries, requires handling of specific methodological issues (e.g. potential bias and confounding variables), as well as guarantee of scientific integrity (e.g. the exclusion of conflict of interest from sponsors).

Outcome measures
Choice of endpoints
The choice of the primary endpoint may pose considerable problems in the design of clinical studies in orphan diseases. In some cases extra time is needed to accumulate sufficient patient numbers to identify the 'most appropriate' clinical endpoint. Equally, as most orphan diseases have an impact on various organ systems, there may be multiple valid clinical endpoints and thus difficulty in choosing a

single primary endpoint, contributing to the statistical power problem. This clinical issue 'outcomes measures' indicates the value of using generic endpoints such as QALYs.

The Guideline for Clinical Trials in small populations[4] states that the usual approach of pre-specifying the primary endpoint may be too conservative and more knowledge may be gained from collecting all sensible/possible endpoints and then presenting all the data in the final study report. Still, every effort should be made to identify an appropriate hierarchy in the endpoints. In the European regulatory procedure, marketing-authorisation may be regarded grantable by the EMA if, collectively, the data look compelling.

Effect size and responder definition

Aside from issues of selecting the appropriate primary endpoint, heterogeneity in the study population may have a significant bearing on the size of the treatment effect estimated. The assumption here is that the new treatments only reduce disease progression, but that they cannot restore irreversible dysfunction. Often the relevance of a gain in reduction of disease progression may be based on an average improvement in the study population. When orphan medicinal products are granted marketing-authorisation based on phase 2 rather than on phase 3 studies, a clear understanding of which type of patient benefits most is still lacking. For example, if all patients are presymptomatic, or only mildly affected, there might be a high potential for slowing down disease progression. If there is a mixed population with patients already in a severe health state, the average potential for slowing down disease progression may be limited. When for example 50% reduction of disease progression is regarded as a relevant reduction, the overall reduction measured can be less or even close to zero because of limited effect in the more severe patients, although the relevant reduction is shown in the 10% of patients in the early mild state of the disease. The variance in outcomes means that the average reported outcome should be considered carefully.

Another complicating issue is that multiple systems in the body can be affected, which may lead to a high variance in efficacy, contributing to the statistical power problem. When some patients have respiratory symptoms, and other patients have muscle weakness, one overall efficacy scale may not be sufficient to measure the full clinical benefit of a new orphan medicinal product. In addition, the potential for different levels of improvement in different subpopulations may lead to further power problems for efficacy. This is the reason economists prefer the use of a generic, multidimensional measure of health status defined by clinicians and patients, although this may lead also to power constraints in RCT's, which are designed for assessment of the standard criterion effectiveness. However, in cost-effectiveness modelling uncertainty is dealt with by other means, so the importance of deterministic statistical significant differences is reduced.

If orphan medicinal products may be rejected on the basis of an average outcome, stratification of the study population offers the possibility of identifying subgroups of patients, who do benefit. This provides the basis for optimising the use of orphan medicinal products in those patients that benefit and subsequently for improving the cost-effectiveness ratio and affordability of expensive orphan medicinal products. However, as we indicated before, a main constraint of this subpopulation approach is that sample size constraints and statistical power problems may become even larger. Therefore, there is a great need for more research in the area of biomarkers and/or clinical efficacy markers in order to select the responders. Pharmacology studies may help identify sources of heterogeneity in patients. Non-clinical pharmacology may sometimes be helpful, especially in conditions affecting very few patients.

Once the outcome measures are defined, the time horizon in which clinical benefits are achieved might create difficulties. The potential long-term effect of the orphan medicinal product in for example disease progression, requires a long-term follow-up with sufficient patients. However, this is an issue for all products and can be dealt with through modelling, as currently happens with non-orphan products.

Patient benefits
As outlined above, there will be great difficulty in defining and measuring clinical outcome and observing a change which is significant by conventional standards. Therefore, if it is not possible to tailor the clinical outcome in patients, it may be more important to make direct observations of the benefits as perceived by patients, which can then be reflected in generic health status and quality of life changes.

Example[6 7]
What is a relevant improvement in the parameter "6 minute walking test" in Pompe disease? An improvement in the patient's ability to walk an additional 20 meters could potentially mean the difference between being able to walk to the car and go to work rather than being compelled to stay at home.

The question is who decides what is a relevant change in outcomes for patients which society might consider worth funding? It could be one of the patient, the family, the treating doctor, clinical experts, regulatory authorities, HTA bodies, the public, or all of them? For the definition of a relevant gain we need to know the social significance of clinical changes observed by the experts. Therefore, a relevant gain is best determined by physicians, patients and their carers. Ultimately, a relevant gain for society is one which makes the treatment cost-effective using conventional cost-effectiveness thresholds, and thus must be measured in QALYs. Generic health status measurement instruments can be used to facilitate this in trials, but clinicians will require proxy clinical measures for use in patient management to indicate the likelihood of sufficient QALY gains.

It is necessary that effort is put into defining the relevant gain for outcome measures used in orphan diseases. Determination of a relevant gain for outcome measures provides the basis for optimising the use of orphan medicinal products in those patients that benefit to an extent that is defined as relevant. Subsequently, professionals can use the pre-specified relevance level of gain for decision-making on starting, continuation or stopping therapy and for designing the treatment protocol and treatment guidelines. The advantage of definition of 'relevant gain' is that it gives clear treatment guidance on which patients to treat and to define start- and stopping criteria, leaving less room for interpretation by the treating physician. This can aid resistance to pressure from patient relatives on the one hand and payers (hospital board, health insurer) on the other hand on whether treatment should be terminated.

Determination of relevant gain is useful in improving the cost-effectiveness ratio in a well-defined and transparent manner, for reducing budget impact and for improving affordability of expensive orphan medicinal products by health authorities.

Table 1: Diagram using statistical significant difference and relevant gain as criteria for assessment of outcomes.

Effectiveness outcomes / Statistical outcomes	Outcome ≥ relevant gain	Outcome < relevant gain
Statistical significant difference	+ +	+ -
No statistical significant difference	- +	- -

Health-Related Quality of life (HRQoL)

The measurement of statistically significant outcomes and the above-mentioned issues related to small sample size of clinical trials and heterogeneity of the patient population are even more a problem for the measurement of HRQoL using disease specific instruments. Given the small sample size, it is particularly important to minimise avoidable missing data that are self-reported by the patient. Standardised administration guidelines are necessary to reduce the variance due to measurement error. HRQoL is a unique patient-reported outcome measure that gives the voice to the patient on the impact of disease and treatment on their daily life.[8]

The QoL of the parents and siblings may also be heavily affected in children with orphan disease, which suggests that ideally this should be included in the analysis as well, although this is not specific for orphan disease. This is not systematically done in current cost-effectiveness analysis, but it is receiving more attention in some countries, so what is important in orphan diseases could become part of the general cost-effectiveness approach in future.

The EQ-5D (or equivalent generic) outcomes are important in the decision-making process, but disease-specific QoL outcomes can be relevant as well. If there is a correlation between the QALY gains and improvement in disease specific measures, it gives decision-makers more confidence that the effects of the treatment are producing the QALY gains. Therefore, a generic preference measure (EQ-5D and SF-6D) plus an orphan disease specific questionnaire can be recommended (if it is possible to construct such an instrument given the problems of diversity outlined above). Indeed, the recent European network for Health Technology Assessment (EUnetHTA) guideline (Feb 2013)[9] on health-related quality of life and utility measures to be used for relative effectiveness assessment of medicinal products also recommends the use of both a generic utility and disease-specific instrument. Their recommendation applies either when the purpose of a relative effectiveness assessment is to inform patients and health care professionals about the HRQoL benefit of an intervention as compared to its comparator, or when the purpose is to inform health care policy makers about the relative value of a product.

HRQoL is multi-dimensional, hence the importance of using multi-dimensional generic preference-based measures of health, such as the EQ-5D and SF-6D. As stated above, instead of focussing only on achieving a relevant gain in clinical outcome measures, the achievement of the pre-defined relevant gain in HRQoL should also be considered in determining the appropriateness of the treatment.

In the HTA report for a treatment for Fabry disease, it is reported that the SF-36-score ranges from 0 to 100, and that a difference of 3 to 5 points is considered a clinically relevant gain.[10]
From a societal point of view this difference might be seen as meaningless, unless it can be shown that it leads to a change in health status using the preference-based SF-6D.

1.4   Relevant *economic* issues for orphan diseases

In this section, the most relevant economic issues for orphan medicinal products and diseases will be mentioned and described. Although these economic issues are not all unique to orphan diseases, and some may be relevant for other disease areas as well, it is important to consider these issues in the application of cost-effectiveness evaluation of orphan medicinal products as part of the decision-making process for coverage.

Opportunity costs
A major issue for cost-effectiveness evaluation is getting good quality costing data (resource use data). Each country has its own rules for costing depending on the decision-making perspective. Ideally the cost assessment is based on the so-called opportunity costs of utilising a technology, which are defined as the benefits to be gained from the best alternative use of the resources. In a system of perfectly competitive markets, price reflects value, and even though health care markets are imperfect, the market price is routinely used to measure the costs of health care goods and services. If a societal perspective is adopted prices paid are an approximate measure of opportunity costs, but with a health system perspective, prices paid reflect opportunity costs to the health sector. The methodological issues for the cost assessment in orphan disease are not much different from other diseases.

Direct versus indirect costs
In a cost-effectiveness analysis, measuring the extent of costs is an important step. Costs can be broadly divided into two discrete resource categories: direct costs and indirect costs (often labelled productivity costs). Direct costs reflect the monetary burden of the medical care and non-medical care expenditures made in response to a disease. The cost of medicinal products is one type of direct medical cost. Other types of direct medical cost include cost of hospitalisations, cost of physician visits, cost of tests and procedures, and cost of durable medical equipment. Direct non-medical costs include the costs to patients, such as travel to obtain treatment, and the value of time spent by family and other non-professional care-givers in supporting patients. The main indirect cost is lost productivity from the patient being absent from work as a result of morbidity or premature mortality induced by the disease or its treatment. Productivity costs are only relevant when studies are conducted from a true societal perspective.
For collection of data on utilisation, a variety of approaches exists. These include subject interviews, subject surveys, provider surveys, medical record reviews, health care utilisation diaries, and insurance claims data reviews.[11] In orphan diseases, primary cost data collection may be necessary due to the lack of historical databases containing information on the health care utilisation costs of

current standard care. The prospective collection of cost data might well be achievable, but in some cases no effort is put into collecting direct medical and non-medical costs as well as indirect costs, if the medicinal product costs are so high that other costs would only contribute a negligible part of the total costs. It is a strong assumption to ignore the non-product costs a priori. It is only possible to draw this conclusion after a sensitivity analysis so full costing is advisable.

The societal perspective is important for orphan disease, as well as for other diseases like depression, and Alzheimer's disease. Consequently a study from a third party payer's perspective, which only includes direct medical costs, would underestimate the true value of a new treatment in orphan disease, particularly one affecting children, in which it is also important to consider long-term costs. Short-term costs may include: time spent by the child undertaking treatment, with loss of educational and leisure activities; productivity losses for parents caring for the child (for example, one parent may decide to stop working); as well as the costs of treatment.
 Long-term indirect costs may be due to productivity loss as a result of 1) the inability of the future adult to work or 2) restricted employment opportunities because of reduced educational attainment or physical disabilities. Therefore the cost assessment in orphan disease should include not only direct medical and non-medical costs, but also the indirect costs for the patient and caregivers in the short and long term. Registries and observational studies may be potential sources for the collection of actual medical costs in daily practice, as well as direct non-medical costs and indirect costs.

Example
A review of an assessment report from the Netherlands for enzyme therapy for Fabry disease leads to following observations[10]: Direct medical and indirect costs were determined on the basis of questionnaires. The direct non-medical costs (e.g. travel and informal care costs) were not taken into account because of their minimal contribution to the total cost. The assumption was made that costs incurred per average patient per cycle per health condition would be the same for treated and untreated patients, and for women and men, with the exception of the orphan medicinal product costs. Given the high cost of medical treatment in this case this assumption would be defendable. Furthermore, potential long-term cost savings by the orphan medicinal product were not included either.
This example shows that the collection of medical costs, direct non-medical costs and indirect costs in prospective studies requires a more detailed and accurate approach in those cases where these costs are expected to affect the cost-effectiveness outcomes, minimising the need for assumptions and missing data.

1.5 Discussion

Discussion
In this chapter, we listed the most relevant clinical and economic issues that are perceived to complicate the cost-effectiveness evaluation of orphan diseases and orphan medicinal products: sample size and heterogeneity, stratification, outcome measures, patient benefits, direct versus indirect costs. Most of the above mentioned issues all originate from "small numbers", which is the leading issue. It is important to consider that each separate issue, or the combination of issues, might

not be unique to orphan diseases. We are aware of other diseases, for example certain cancers and multiple sclerosis, whose evaluation might also be hampered by the same (combination of) issues. For example, in multiple sclerosis multiple systems in the body are affected, which may lead to a high variance in efficacy outcomes and variation in the impact of interferon treatment on disease progression. The requirement for a long-term follow-up period also applies to multiple sclerosis. Also various new expensive cancer drugs which are only effective in specific patient subgroups, exhibit the characteristics of orphan products. On the other hand, sample size constraints are not so much an issue in multiple sclerosis and oncology, and comparative therapeutic alternatives are already available, allowing appropriate RCT's to be performed. Therefore, medicinal products, whether they are orphan or non-orphan, that are hampered by limitations to the clinical data at the time of launch, may require a different approach in HTA assessment procedures. For the purpose of this report we used the orphan medicinal products as an example to elaborate on the clinical and economic issues.

If the total set of clinical data (effectiveness) of an orphan medicinal product at time of launch shows particular limitations and if economic issues also exist (for example, the often high price of the orphan medicinal product), how does all this affect the cost-effectiveness evaluation? From what we have seen in practice, the limitations of the clinical data combined with the relatively high price of the product result in high to extremely high ICERs not meeting any national threshold for cost-effectiveness, or result in a large confidence interval around the ICER due to the uncertainty over the clinical outcomes. As a consequence, the orphan medicinal product risks being rejected for coverage.

Rationale of this project
This situation needs a solution if we wish to encourage the development of new products to treat orphan diseases. Theoretically, two possible solutions are available: 1) circumvent the standard assessment criterion - cost-effectiveness- by adding other criteria reflecting social values, or 2) keep the standard assessment criterion - cost-effectiveness- and put effort into optimisation of the inputs for the cost-effectiveness evaluation, for example by limiting the use of these medicinal products to those patients who show maximum benefit and in such a way (containing stringent start and stopping rules) that the associated cost-effectiveness is still within reasonable limits or at least drops substantially.

Before elaborating on these two possible solutions, the role of cost-effectiveness evaluation in coverage decisions for orphan medicinal products remains to be discussed. We address this in the next chapter.

2.   Role of cost-effectiveness evaluation and other economic instruments for coverage decisions

2.1   Introduction

The decision of health authorities on coverage of a medicinal product under the national health insurance package is based on the value for money of a new medicinal product. Health authorities will make a trade-off between the incremental, clinical benefit and the premium price of the new medicinal product versus standard therapy.

A health technology assessment (HTA) would not be complete without a cost-effectiveness evaluation. Yet, the relationship between cost-effectiveness and health authorities' decisions for covering medicinal products in the health insurance package is not always clear and consistent. A "grey" area in decisions using cost-effectiveness data has been noted, with technologies with favourable cost-effectiveness not covered (e.g. Viagra for erectile dysfunction) and technologies with far less favourable cost-effectiveness covered (e.g. oncology therapies, liver transplant). Authorities are hesitant to disclose the rationale for their decisions for both political and technical reasons. Moreover, transparency is hampered if a manufacturer submitting cost-effectiveness data claims confidentiality of pricing and manufacturing costs to prevent competitors gaining access to this information. Although many European countries have official requirements for submitting cost-effectiveness data for coverage submissions, authorities may use an unofficial kind of multi-criteria decision-making framework without the obligation to make public their judgement on these data or its weight in the decision-making process.

In addition there are no thresholds for approving coverage for each of the various data requirements. Health economists may advise on the interpretation of a cost per QALY, but the decision about how much society will pay for increased effectiveness is political.

Role of cost-effectiveness in current coverage decisions on orphan medicinal products
The lack of a completely clear relationship between cost-effectiveness and health authorities' decision-making processes on coverage may create uncertainty in those submitting applications for coverage. Is it essential to demonstrate cost-effectiveness, or can special categories of product, such as orphan medicinal products, succeed in other ways?

> In this chapter the question is addressed: what weight should be attached to the cost-effectiveness evaluation amongst the set of standardised criteria applied in coverage decision-making in Europe, in particular for orphan medicinal products.

Recent examples illustrate the need for transparency with regard to the role and weighting of cost-effectiveness and decision-making in coverage procedures for orphan medicinal products.

In the UK, the National Institute for Health and Care Excellence (NICE) has calculated ICERs for orphan medicinal products of between £200,000 - £300,000 per QALY gained (€238,000-€357,000).[12] For example, for the enzyme replacement therapy, agalsidase beta (Fabrazyme), NICE has calculated an ICER of £203,009 per QALY gained (€241,580). In the Netherlands, the health authorities have calculated an ICER of €3,3 million per QALY gained for agalsidase beta (Fabrazyme) and also for algasidase alpha (Replagal).[6] Subsequently they judged the cost-effectiveness as insufficient using standard coverage methods.

The huge difference in ICER between the UK (€241,580) and the Netherlands (€3,3 million), raises questions about the methods used, as it cannot be explained by differences in discounting, treatment patterns, and unit costs. Such large differences are most probably explained by differences in the patient groups assessed (indication) and the valuation of the health benefits. The cost-effectiveness ratio of enzyme replacement therapy extends far beyond the bandwidth of the Dutch Council for Health[3], which ranges from €20,000 per QALY gained in case of a low disease burden (0.1) to €80,000 per QALY gained in case of a high disease burden in The Netherlands. The National Health Insurance Board (Dutch College voor zorgverzekeringen, CVZ) admitted that "For orphan medicinal products, this bandwidth cannot be used as a hard decision criterion, as these products may be more expensive than usually. With a higher price and a longer payback period manufacturers have incentives to develop orphan medicinal products". A reason for this milder attitude probably was that orphan medicinal products are often used in situations where other factors favorable to positive coverage decisions are present. There may be a high disease burden on individual patients and small numbers may mean that there are "identifiable victims" (if care is denied) and a small budget impact. However, CVZ advised the Minister of Health to exclude these products from coverage under the health insurance package due to the low cost-effectiveness.[7 10] Then, under heavy societal pressure, the CVZ adjusted their advice into preliminary continuation of the already existing conditional coverage since the time of launch of these products in 2001, while at the same time working on alternative ways to optimise use. They tried to reduce the cost by requiring use to be supervised by specialists, negotiating with manufacturers about the price, and by seeking to transfer these costly products to a special separately-financed fund for orphan medicinal products. The Minister of Health agreed to the first two suggestions but denied the last one, referring to the general legal basis of coverage of medicinal products that also applies to orphan medicinal products.[13] A financial agreement between the Minister of Health and manufacturer was achieved shortly afterwards.[14] This situation illustrates precisely the lack of consistency in the relationship between cost-effectiveness and health authorities' decisions on covering medicinal products in the health insurance package. Furthermore, this gives the impression that coverage decisions are made on an ad hoc basis, leaving stakeholders with questions about the formal role and weighting of the standard assessment criteria.

Reviews of orphan medicinal products by the Scottish Medicines Consortium (SMC) also followed the standard methods of assessment, resulting in expected rejections for coverage in the NHS in Scotland. Aldurazyme for mucopolysaccharidosis I (MPS-I) was turned down because of "failure to submit evidence to support being cost-effective". Myozyme for the treatment of Pompe disease was rejected because "the economic case was not made" with cost per QALY gained ranging from £224,000 (€266,560) in infants to £819,000 (€974,610) in adults.[15] The SMC did note that they might consider additional factors in reviewing orphan medicinal products, but only mentioned clinical

parameters (population, disease progression, and others). There was no indication of how other data pertaining to the social value of medicines were considered in their decisions.

Discussion

These examples from the UK and the Netherlands reflect the existence of the aforementioned clinical and economic issues, and the subsequent rejection of orphan medicinal products for coverage due to (extremely) high ICER's. Furthermore, they illustrate that straight forward application of the common thresholds for standard cost-effectiveness outcomes to medicinal products with limitation of the clinical data at the time of launch, may not always be acceptable to decision-makers. It seems that health authorities are also aware that such medicinal products require a different and more flexible approach. Currently, several European countries have a framework for this under development. Within these frameworks, there seems to be a tendency towards a lower weighting of the incremental cost-effectiveness outcome in the overall coverage decision for orphan medicinal products and the addition of criteria other than maximising population health. In section 2.2 we review two examples of recently published frameworks for coverage decisions on orphan medicinal products.

Another important trend, working in the opposite direction, is the development and implementation by health authorities of additional policy tools to control the use of products which have passed the traditional coverage assessment criteria set. In section 2.3 we provide examples of these policy tools with which health authorities are able to restrict the use of, and the total budget spent on, products to a level below that indicated by the application of cost-effectiveness criteria. The difference from the presented frameworks (discussed in section 2.2) is that these policy tools are additional to the standard assessment criteria set, whereas the frameworks are proposed for application in place of the standard assessment criteria set.

2.2   Current coverage frameworks in Europe

In this section we review two recently published proposals for a framework for the valuation of orphan medicinal products for coverage under the health insurance package.

I. The first proposal for a framework for the assessment of orphan medicinal products is described in Hughes-Wilson et al. (2012).[16] They argue that to date, orphan medicinal products have only accounted for a small percentage of the overall medicinal product budget. They claim that these orphan medicinal products accounted for just 1.9% of medicinal product expenditure in Belgium in 2008[17], rising to around 2% in 2009[18], while in France and the Netherlands, they accounted for just 0.7% and 1% respectively of national medicinal product budgets in 2004.[19] They comment that fears that growth in orphan medicinal product expenditure will lead to unsustainable cost escalation, may not be justified.[20] They identify two main criticisms of behaviors in the orphan medicinal product field: the high prices of orphan medicinal products and their inability to meet standard cost-effectiveness thresholds. They state that it has been acknowledged that standard methodologies for Health Technology Assessments (HTA) will need to be tailored to take into account the specificities of orphan medicinal products given that the higher price claimed by orphan medicinal products are unlikely to meet current cost-effectiveness thresholds. The authors propose the development of a

new assessment system based on several additional evaluation criteria, which would serve as a tool to evaluate each new orphan medicinal product at the time of coverage assessment. They suggest that such a system should include criteria such as rarity, disease severity, the availability of other alternatives (level of unmet medical need), the level of impact on the condition that the new treatment offers, whether the product can be used in one or more indications, the level of research undertaken by the developer, together with other factors, such as manufacturing complexity and follow-up measures required by regulatory or other authorities (see Table 2). Each individual government would then decide on the weighting attributed to each of the criteria in question, based on what each individual society values most, acquired through a survey of public preferences. For the purpose of performing such a survey, the authors propose a discrete choice experiment, involving members of society, to elicit and quantify the relative importance of each criterion in the evaluation of orphan medicinal products in a multi-criteria decision analysis (MCDA) process. Each criterion needs to be measurable, so that the degree to which an orphan medicinal product attains the criterion can be assessed by the individual government. The scores of an orphan medicinal product on the different criteria are then aggregated with a view to calculating an overall performance score. Importantly, this would allow governments to value an orphan medicinal product that fulfilled all the criteria very differently from one that only met some of them.

Table 2: Proposed criteria for evaluation of orphan medicinal products and corresponding potential parameters.[16]

Criteria	Price Differential		
	Lower	Medium	Higher
Rarity	1:2,000 – 1:20,000 or COMP figures > 3 in 10,000 (11%)	1:20,000 – 1:2,000,000 or COMP figures 1-3 in 10,000 (51%)	Less than 1:200,000 or COMP figures less than 1 in 10,000 (3,8%)
Level of research undertaken	Literature review	Building on previous existing knowledge	"Blue-sky" – starting research & development programme in an unknown area
Level of uncertainty of effectiveness	Immature, but promising data	Appropriate surrogate end-points	Robust clinical end-points
Manufacturing complexity	Not complex – small molecule / classic galenic form	Moderately complex	Highly complex biological and galenic form
Follow up measures (additional benefits and associated costs)	Moderate to none	Designed to answer specific, defined, delineated question	Safety and efficacy studies + size and duration of study
Characteristics without direct cost impact			
Disease severity	Morbidity	Mortality / severe invalidity in adulthood	Mortality / severe invalidity as infant
Available alternatives / unmet medical need	Alternatives with similar characteristics	Alternatives – but offering strong innovation to the disease treatment	No alternative
Level of impact on condition / disease modification	Low	Medium	Strong
Use in unique indication or not	Existing orphan or non-orphan indications for the same molecule*	Potential for multiple indications	Unique indication – no other use possible

*Note: Another element could be the total revenues in the context of multiple indications for the same molecule owned by the same company.

II. Sussex et al. (2013) identify a number of possible attributes to include in an orphan medicinal product value framework, using a MCDA process for determining the relative importance of the attributes, and testing whether an MCDA approach can practically support decision-making.[21] The project included literature searches on the natural history and burden of forty rare diseases and on how payers assessed treatment value. Also three workshops were held: the first workshop held was with GlaxoSmithKline (funders of the research project) managers working on orphan medicinal products, the second with EU clinical and health economics experts, and the third with representatives of rare diseases patient groups in the European Union. In each workshop participants refined the attributes, assigned weights to them, scored two case studies with orphan medicinal products in terms of those attributes, and tested the sensitivity of the overall ratings to changes in weights and scores. This process yielded eight non-monetary attributes: four concerning the disease being treated and four concerning the treatment itself. The eight attributes are as follows:
1. Availability of effective treatment options / best supportive care in the absence of the new medicine
2. Disease survival prognosis with current standard of care
3. Disease morbidity and patient clinical disability with current standard of care
4. Social impact of the disease on patients' and carers' daily lives with current standard of care

5. Treatment innovation, defined as the scientific advance of the new treatment together with contribution to patient outcome
6. Evidence of treatment clinical efficacy and patient clinical outcome
7. Treatment safety
8. Social impact of the treatment on patients' and carers' daily lives

The study took a societal perspective when establishing the full set of value attributes, while recognising that payers and HTA bodies in some countries currently take narrower perspectives limited to measures of clinical effectiveness or health gain. Similarly, they asked participants when determining attributes' relative weights to take into account the interests of all relevant stakeholders including patients, their families and carers, payers and the national economy.

Participants unanimously confirmed that the MCDA approach provided clarity, logic and transparency. The authors conclude that given the intrinsically complex nature of the rare diseases and orphan medicinal products' environment, an MCDA approach for rare disease treatment value assessment has the merit of ensuring shared understanding of the elements of value as well as a clear articulation of trade-offs between those elements. Furthermore, the authors state that the MCDA approach and resultant value framework are complementary with current EU policies on orphan medicinal products. According to the authors, the orphan medicinal product value framework derived via MCDA offers a possible construct for more comprehensive guidance to HTA and coverage decision-making.

Discussion
Although the MCDA approach seems to support the integration of social values, whereas the cost-effectiveness evaluation only considers efficiency (based on a particular set of assumptions about social values), we put question marks on the potential of the MCDA approach to provide helpful answers. One particular problem is identifying a representative group (to engage in the exercise) so that the outcomes represent real societal preferences. Given that cost containment in health care is so important these days, a more important problem with this approach is, that there is a risk that all these exercises come up with decision criteria that (when applied in actual coverage policies) are not sustainable in the long run. In these circumstances a series of different trade-offs between health maximisation and the other elements of social value will need to be identified if the budget constraint is to be met.

2.3   Emerging instruments for coverage policies

Recently, health authorities have started to develop additional policy tools because the traditional instruments, cost-effectiveness and budget impact analysis, are not always considered to fully represent social preferences and are not considered to contain health care budgets sufficiently. At best these tools can help to achieve cost-effective outcomes by better targeting treatments; at worst they can be crude attempts to cut costs regardless of efficiency or patient benefit. Health authorities are currently developing and implementing these new policy measures, and no final judgment can yet be made on their success. These new emerging instruments are described in this section with the

aim of further understanding whether they could be helpful in coverage decision-making models for orphan medicinal products.

2.3.1 Prescription restrictions

<u>The concept</u>

The development of prescription restrictions is a policy tool to control the budget impact of medicinal products. While traditionally coverage decisions applied to the officially registered indication, authorities have recently been imposing restrictions on the coverage for medicinal products. These restrictions may include: following a treatment protocol, limiting the number of prescribing physicians or limiting the range of indications (United Kingdom, Germany and France). Arbitrary limits on the use of medicinal products purely to meet financial targets clearly do not lead to an efficient use of health care resources. However, if the restrictions are based on evidence of how much use is cost-effective, they can control expenditure and lead to an efficient use of resources. For example, in the United Kingdom the NICE is producing clinical guidance incorporating clinical and economic evidence, which includes specific advice on targeting medicinal products, although it is not legally binding. Through the creation of treatment guidelines for certain conditions, the healthcare authorities are seeking to exercise control over physicians' activities including prescribing. These guidelines are distinguished from clinical guidelines from specialty associations or opinion leaders.[22] The primary goal of the latter is to maximise the health outcomes of a treatment based on the clinical evidence (efficacy, safety) and expert opinion, without regard for the opportunity cost of such treatment. In contrast, the primary goal of NICE guidance is to identify the most efficient way to treat a patient group using systematic review of the clinical evidence and cost-effectiveness analysis. Even cost-effective treatments may not be affordable in the short-term, so a budget impact analysis (BIA) may be needed. This should cover the whole of the impact on health care expenditures, not just the cost of medicinal products, but in many countries there is a tendency to focus on control of the medicinal product budget per se.

<u>Application to orphan medicinal products</u>

The creation of treatment guidelines might be relevant for orphan diseases, incorporating guidance on the use of orphan medicinal products. Any prescription rules should be based on clinical and economic evidence instead of administrative criteria, such as, 'prescription by specialists only'. The development of prescription rules is closely related to another instrument, discussed in the next section: 'stratified medicine'.

2.3.2 Stratified Medicine (SM)

<u>The concept</u>

Stratified Medicine (SM) represents a novel approach to increase medicinal product research & development efficiency and to provide improved medical outcomes for the patient and the health care system. This approach of proactively testing and selecting populations for specific treatments aims at ensuring a differential patient response by either increased efficacy and/or reduced toxicity within the selected populations, but at the same time it reduces the addressable patient population for the treatment to the selected subset. The link between the clinical biomarker and the preventive and curative therapy provides new opportunities for value creation, it offers the potential to change

well-established practices for physicians and it strengthens the value proposition for coverage decisions of authorities.

Stratified Medicine, as opposed to empirical medicine, is the practice of using biomarkers or diagnostic tests to associate a patient with a specific therapy whereas by contrast, in empirical medicine all patients would receive the same treatment (Figure 1).[23] Trusheim et al. (2007)[23] consider such tests can be based on gene expression patterns, individual proteins, proteomic patterns, metabolomics, histology, imaging, physicians´ clinical observations and even self-reported patient surveys. In other words, they define a clinical biomarker not by its technology or biological basis, but rather by its reliable, predictive statistical correlation with differential prognosis or different patient responses. Often, tests are developed to prove clinical validity (sensitivity and specificity of the test) without evaluating clinical utility. Currently, many of the available tests in clinical practice do not have sufficient evidence without evaluating clinical utility, which makes it difficult to demonstrate the value of diagnostics. At the extreme of SM are 'individualised' medicines, which vary inherently for each patient such as cancer vaccines that are based on a particular patient's tumour, representing one end of a patient continuum (Figure 2).

As more SM products reach the market, public and private third party payers in the US and Europe will demand better evidence to make more informed coverage and resource allocation decisions at both the local and national levels. Generating high quality clinical and health economic evidence will provide the confidence that enables payers more rapidly to adopt tests and to align physicians' incentives with patient care and outcome rather than procedure. At the same time, payer decision-making may need to become flexible enough to allow for short-term inefficiencies in order to understand and benefit from long-term value. While the need for market access of diagnostic-based therapies is not questioned, the framework for access, while setting the right incentives and appropriate alignment of stakeholders for further innovations, still needs careful development.

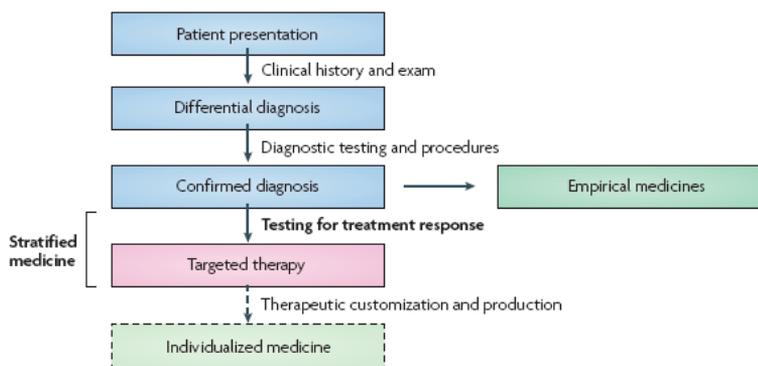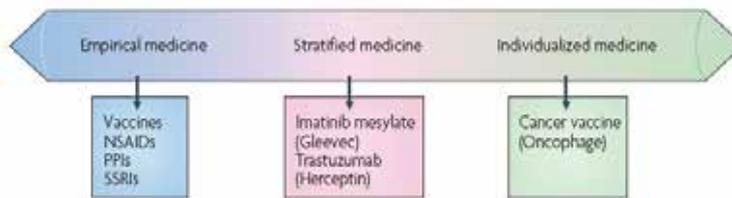**Figure 1: Stratified Medicine in the clinical context.[23]**

Figure 2. Stratified medicine diagram (source: Trusheim et al. The patient therapeutic continuum).[23]

Empirical medicine | Stratified medicine | Individualized medicine

Vaccines
NSAIDs
PPIs
SSRIs

Imatinib mesylate
(Gleevec)
Trastuzumab
(Herceptin)

Cancer vaccine
(Oncophage)

Empirical medicine is at the other end of this continuum where some agents work for almost all relevant patients such as non-steroidal anti-inflammatory drug (NSAIDs). In between lies the field of stratified medicine, in which a patient can be found to be similar to a cohort that has historically shown a differential therapeutic response to a particular therapy using a clinical biomarker that has been correlated to that differential response. Often-named examples are: HER2/neu – Herceptin, KRAS/EGFR –Vectibix & Erbitux, predictive for efficacy; UGT1A1 /Irinotecan, HLA-5701/ Abacavir (HIV), VKORc1/CYP2C9/ Warfarin predictive for safety; Oncotype DX and MammaPrint prognostic for adjuvant chemotherapy.

To foster a broader coverage of SM approaches within the healthcare systems, there is a need for a more consistent process for conducting HTA´s in Europe and in the US. As long as no commonly accepted standards on how to evaluate diagnostic-based therapies have been transparently established by leading HTA agencies and payers, a population wide use of such interventions is difficult to reach. In Europe for instance, until recently, there was no single leading HTA group responsible for evaluation of more complex diagnostic testing. In the US several health plans have expressed the need for more guidance on coverage issues with new diagnostic technologies. More comprehensive HTA's, which take into account the most reliable, available evidence and are tailored to the specificities of SM interventions, will better identify the cost-effectiveness of such applications. At the same time, a more holistic approach to health care funding is required in order to realise the full clinical and health economic benefits of SM interventions. The existing modelling techniques are appropriate and can be applied for cost-effectiveness models in SM. The inclusion of sensitivity/specificity and especially false negatives and false positives, requires additional structural complexity in order to make the link between the test and the treatment choice for medication. Another issue is the gaps in information, especially for stand-alone tests. Information on treatment patterns, costs and outcomes, are often lacking, especially for false positive and false negative patients. Finally, statistical methods are required in order to extrapolate the short-term sensitivity/specificity data to long-term cost-effectiveness. In the case of genomic technologies, cost-effectiveness analyses need to incorporate post-marketing innovation and learning by-using sensitivity analyses in a more systematic manner.

To realise the promise of SM approaches, there is a need to perform economic evaluations which take into consideration the full impact of using such an intervention on the whole treatment pathway of patients including disease prevention. The cost and effectiveness of the different therapeutic options in the treatments pathway change when patients are being identified as a responder or non-responder to a specific therapeutic option, and hence demand a comprehensive value assessment with a system-wide perspective taking into consideration the costs and benefits of having less adverse events and more therapies targeted towards those who benefit most.

Medical diagnostics is fundamentally about identifying subgroups of patients, for whom the expected benefits of treatments would outweigh the risks and consequently may increase adoption and compliance by responders. In clinical practice, much depends of the reliability and predictability of diagnostic tests in order to show clinical utility to change patient management.[24] Currently, many orphan diseases lack available tests in clinical practice that have sufficient predictive evidence. For example, in the non-classical subpopulation with Pompe disease, a positive test may not really predict that a patient will ever develop Pompe disease. Therefore treatment based on a positive test may be redundant in a proportion of patients and may lead to an unnecessarily high impact on the medicinal product budget. However, if you do not treat the "patient", and he/she develops Pompe disease, the opportunity of treatment in an earlier stage may be lost. Therefore the decision for "treatment" or "no treatment" may also lead to dilemma that spending money on orphan treatments with no proven benefit is denying benefits to patients with other conditions and evidence of benefit. Another example is illustrated by the orphan disease 'severe methylenetetrahydrofolate reductase (MTHFR) deficiency'. Early treatment of severe MTHFR with betaine is associated with reduced mortality and normal psychomotor development, whereas in case of delayed treatment brain damage has already occurred, and patients exhibit abnormal psychomotor development despite treatment. Methods for high-throughput detection of homocysteine have recently become available for the pivotal, timely detection of this disorder.[25]

Therefore more research is required in diagnostic testing and biomarkers corresponding with the principles of stratified medicine. This will allow a better identification of persons at risk. Also, case-control, observational and patient cohort studies are used to determine the clinical value of biomarker-based diagnostics when randomised control trials (RCT) are not feasible or are too expensive.[26] The design of prospective outcomes research studies in orphan disease should include components for prospective outcomes research linked with SM.

2.3.3 Value based pricing (VBP)

The concept

Recently, novel payment approaches, such as risk sharing and conditional coverage agreements with third party payers, have been explored to overcome the tension between funding new but expensive health technologies and obtaining value for money where traditional coverage is not deemed appropriate. These arrangements between a manufacturer and payer/provider can use a variety of mechanisms (e.g. pay-for-performance, value-based purchasing) to address uncertainty about the real performance of technologies in daily practice. While in the US such outcome-based approaches are increasingly being adopted for high priced stand-alone diagnostic tests, in Europe, with its primarily centralised health care systems, there is so far little experience of such arrangements for high priced medications or diagnostics known in the public sector.

One exception to this is the UK, which has proposed the introduction of a system of linking the price of new medicinal products more closely to their demonstrated social value. The proposals, described as VBP, build on the role of the NICE in assessing the cost-effectiveness of treatments, by including factors other than the quality-adjusted life year (QALY) in the appraisal of benefits. Additional factors include burden of disease (defined as a combination of unmet need and severity), health effects not measured by the QALY, and the wider social impact on, for example, carers and patients' families.

These developments will be introduced in 2014, without major changes to the deliberative decision-making processes used by NICE committees or to the NHS and Social Care perspective used in analyses for NICE.

Within the VBP concept, we can identify:
Elements of value:
1. Health effect is usually the single most important benefit of health technologies. Direct health effects can be measured using indicators of efficacy or effectiveness such as the QALY, which combines changes in the quality and length of life of an intervention.
2. Any cost-offsets within the healthcare system are a second key benefit. Savings to the health care system (offset by the additional cost of using the technology) are usually included in standard cost-effectiveness analyses.

Other elements of value fall into three distinct types:
3. A QALY's "value" to society may be higher or lower depending on who gets it. This might depend on the characteristics of the patients receiving the health gain (for example, age), on the nature of the illness in question, or on the pre-treatment level of health or disability of the patients.[27] The UK's VBP proposals suggest that the value of the QALY should be weighted by disease severity.[28]
4. Elements of benefit to the patient that are not necessarily captured in the QALY (or any other measure of health gain), including:
   a. Health related quality of life aspects not well reflected in a generic measure. For example, vitality is an important aspect of cancer patients' health, but it is not explicitly included in the EQ-5D[28], which is one of the most used health measurement systems. However, it is difficult to believe that the impact of changes in vitality is not picked up by the EQ-5D domains (mobility, usual activities, anxiety and depression and self care).
   b. Health care process related aspects, such as being treated with dignity, at a convenient time and location, and after only a short wait. These may have health consequences, but the preference for them (as reflected in patients' stated preferences, or in political targets; for example, waiting times) goes beyond any health gain. These aspects reflect universal health care issues which have no bearing on the relative merits of different technologies and therefore cannot influence value.
   c. Information for the patient that, for example, enables life style choices to be made independent of any health effects. However, the societal gain from this is less easy to identify and measure.
5. Other costs and benefits beyond those to patients and the NHS, such as the benefits to employers of getting people back to work more quickly and quality of life improvements for carers.

Options for aggregating elements of value into a VBP:
6. Considering each type of benefit in terms of its own "unit of measurement", and applying a set of weights to each benefit type to represent the rates at which different types of benefit may be traded-off with each other, and scores to indicate how well each benefit type is achieved by the medicine in question. This is called a MCDA approach.[21][24]
7. Selecting one principal measure of benefit, the default option being QALY's, as the "numeraire" and then up-rating or down-rating that measure using a series of weights to reflect the magnitudes

of other types of benefit. Another option would be to assess (using stated-preference approaches) how people trade off QALY gains with other value elements such as informational benefits that are independent of health gains.

8. Using a "deliberative process" of the sort used by NICE and other HTA bodies, where considerations other than QALY's are assessed and weighted qualitatively. In most deliberative processes, the relative weights given to the elements of value may remain implicit.

The principal approaches are considered in more detail in Table 3, which highlights some key issues and advantages of each, as well as common challenges.

Table 3: Approaches to the aggregation of overall value; issues and merits of each; and implications for the identification of the value-based price.[21]

	How is value aggregated?	Key issues specific to this approach	Key merits of this approach	Issues common to all approaches
Net benefit	As the sum of the benefits, each assessed in monetary terms.	Challenges estimating the value in monetary terms of each type of value. Allocating a monetary value to health had been always one of the mayor criticisms.	Arguably, a better grounding in economic theory. Facilitates the comparison of value and value for money across health and other sectors. Use of monetary value may resonate better with some (private) payers	A consensus on the perspective (NHS? Government? Societal?) from which value is assessed is required regardless of which approach is used. The metrics by which aspects of value other than health are measured needs to be defined, as a prior step to valuing them.
MCDA	As the sum of the points assigned to each aspect of value.	The cost-effectiveness threshold would need to be reassessed in terms of the cost per incremental *point*.	A pragmatic approach, widely used in the UK public sector. A more transparent (than a weighted QALY of deliberative process alone) means of addressing multiple criteria. MCDA is used in local NHS commissioning: potential to develop a consistent priority setting framework for both new and existing health care technologies.	
Weighted adjusted QALYs	1. By QALYs gained, up-rated by one or multiple weights to represent the magnitudes of other aspects of value; or 2. direct estimation of how people trade off QALY gains with other value elements.	Assumes that all other sources of value are proportional to the number of QALYs gained. There are implications for the threshold. If the value of new technologies is assessed in terms of a range of criteria, then opportunity cost also has to be considered in the same terms, not just QALYs foregone. Even if a simple social weighting or QALYs applied, the opportunity cost will change.	Is it relevant to state here the classic arguments favour of the QALY such: Allows for the comparisons therapeutic areas in the NHS "A QALY is a QALY" argument Well established in the UK within HTA bodies (and academic centers) Understood by health economics community.	
Deliberative process	Weights are assigned by a committee to each relevant aspects of value.	Weights are often implicit, implications for the threshold.	Provides an element of flexibility. Well-recognised approach used by HTA bodies around the World.	

All these issues may be especially relevant and useful for the coverage of expensive medications, including orphan medicinal products. Although they may not ultimately have much impact on decisions, taking them more explicitly into account will improve the transparency of decision-processes.

2.3.4 Outcomes research and temporary coverage

The concept

Clinical trials can have a low external validity because they have strict inclusion and exclusion criteria and treatments are protocol driven, leading to potential overestimation of units of healthcare utilisation. Considering the prospective approach, the concept of validity should be addressed. Internal validity is the extent to which the analytic inference derived from the study sample is correct for the target population. External validity or generalisability is the extent to which the results found in the study sample also apply to the population from which the sample was taken. Randomisation is usually applied to balance confounding variables, and with double-blinding, helps to support internal validity. However, the external validity of RCTs is more questionable. Inclusion criteria for patients and selection of study sites may mean that the sample is not representative of the potential patient population. In addition treatment patterns may be determined by the protocol.[29] Therefore, both clinical and economic outcomes may not be typical and do not correspond to usual practice. Hence it should always be considered that, due to its restriction on external validity, the estimates of efficacy from RCTs may not be representative of the effectiveness of the intervention in the target patient population.

The statistical constraints and limitations of external validity may be alleviated by the use of registries. Registries use large longitudinal, observational studies designed to measure the impact of a particular disease or condition on clinical and patient-specific outcomes, and to document the outcomes associated with different treatments or settings of care. Patients are followed prospectively and data are collected on disease severity and clinical outcomes as reported by clinicians, as well as resource use, functional status and quality of life as reported by the patient. Currently targeted longitudinal observational databases, or patient registries, are being designed, which reflect the current treatment patterns without influencing the treatments and without any interference with real practice (e.g. no randomisation) being fully naturalistic and having a high external validity.

Observational data, however, may also have important limitations through the non-random decisions of clinicians that introduce bias.[30] In observational data, patients would often be treated differently based on the underlying condition.[31] By contrast, in a well-executed RCT, the patients' condition is independent of treatment, and one can therefore reasonably attribute observed effects to the particular variation in treatment being studied. In observational data the treatment is not allocated randomly. As a result, the characteristics of those obtaining the treatment will generally differ from the characteristics of those who do get standard care (the controls). The differences may be observable characteristics such as age, in which case a regression equation can potentially control for them.[32] However new methodologies have been developed, that may reduce the dependence of cost-effectiveness analyses on RCT data. For example, Newhouse presented an econometric technique, instrumental variables, that can be useful in estimating the effectiveness of clinical treatments in situations when a controlled trial cannot be done.[31] This technique relies upon the

existence of one or more variables that induce substantial variation in the treatment variable, but have no direct effect on the outcome variable of interest. The concept was illustrated with an application to aggressive treatment of acute myocardial infarction (AMI) in the elderly, which showed that some of the differences in observed mortality between patients receiving catheterisation and no catheterisation could result from differences in the capabilities of the hospitals and physicians treating the two groups of patients, independently of the procedures given to them. For example hospitals with a catheterisation unit are more likely to have other sophisticated treatments available to their patients, and their physicians and nurses may be more highly trained. These problems in general mean that while observational data is potentially useful, it must be used with caution, as there are a number of examples where observational data have resulted in conclusions about the effectiveness of therapy, which have subsequently been shown to be wrong such as the prophylactic administration of lidocain to patients with AMI.[33]

Application to orphan medicinal products
The temporary acceptance of an orphan medicinal product for coverage might be considered based on the initial clinical data from a pre-launch register. A conditional acceptance would permit initial decision-making on coverage based on the cost-effectiveness of the new medicinal product derived from modelling data at the time of launch, followed by validation through subsequent prospective data collection by means of a larger registry. The registry may yield more statistically solid safety data with higher external validity because of the larger sample size of the registry compared with the initial clinical trial. The larger sample size of the registry may also allow the identification of any type of covariance.

2.4 Discussion

Coverage systems and assessment of the value for money of new medicinal products in many countries are based on methods considered appropriate for traditional medicinal products, which entered the market in the seventies and eighties (e.g. antibiotics, antidepressants) and which all had a relative limited daily price. However, since the late nineties, completely new classes of medicinal products entered the market, which offer treatment for a small number of patients at a high price per patient. Among these products are, for example, biologicals with a completely different mechanism of action, new oncology therapies and orphan medicinal products. These are all highly-priced medicinal products which are likely to be rejected out-of-hand by payers focused on budget impact. Even in those countries with more sophisticated decision-making processes based on HTA, these products will have high ICERs or large confidence intervals around the ICER (usually caused by limitations of the clinical data at the time of launch), and will struggle to gain reimbursement. As a result, some of these products have been rejected for coverage, while others were accepted for coverage, often with little transparency and no clear formalised structure or framework, which justifies the criteria on which the final decision was made.
Due to the lack of a clear relationship between the cost-effectiveness results and health authorities' decisions on coverage, stakeholders have questioned the application of conventional economic evaluation methodologies and have advocated the development of new policy measures. The development of alternative assessment frameworks (such as MCDA, see above), with a lower

weighting of the incremental cost-effectiveness outcome in the overall coverage decision and the addition of other criteria, measuring for example social values, and the new additional policy tools (guidelines, stratified medicine, value based pricing, risk sharing and outcomes research and temporary coverage) are potentially relevant to the coverage of expensive medications, including orphan medicinal products. However, there is insufficient experience of their practical application to make a final judgement.

As currently proposed, most systems for the use of MCDA seem designed to make more products eligible for reimbursement. We therefore question their potential to provide helpful answers and to be sustainable, given the importance of cost containment in health care. With respect to the other policy tools, we see them as dependent on the outputs of a conventional cost-effectiveness evaluation. Once this has been done, additional mechanisms can be applied to ensure that a cost-effective use of resources is achieved; e.g. starting- and stopping rules; treatment guidelines; and risk-sharing agreements.

Conclusions

Based on the above discussions, we do not propose the adoption of a MCDA approach or other policy tools to circumvent a cost-effectiveness evaluation and subsequently to avoid ICER thresholds. Instead of creating new instruments, we suggest that efforts should be directed at optimising the inputs to the cost-effectiveness evaluation, for example by limiting the use of these medicinal products to those patients that show maximum benefit and in such way (containing stringent start and stopping rules) that the associated cost-effectiveness is kept within reasonable limits or at least drops substantially.

In the next chapter we provide guidance on a coverage decision-making procedure, based on cost-effectiveness analysis, for which we have developed an algorithmic approach.

3. Guidance on coverage decision-making in orphan disease: an algorithmic approach

3.1 Introduction

The preceding overview in chapters 1 and 2 of this report showed that there are a number of clinical and economic issues that complicate the application of standard cost-effectiveness evaluation to (orphan) medicinal products, principally because of the limitations to the clinical data available when products are launched. The associated uncertainty may result in a large confidence interval on the ICER, which may, in turn, induce authorities to make a negative coverage decision.

In response to these issues, stakeholders have disputed the appropriateness of the emphasis that is put on the cost-effectiveness criterion in national coverage assessment procedures. They have initiated the development of alternative, value frameworks, containing attributes and agreed weights using a MCDA approach, or policy tools for circumventing cost-effectiveness in coverage procedures for orphan medicinal products. These additional attributes include rarity, disease severity the availability of other alternative treatments, level of unmet medical need, and the level of impact on the condition that the new treatment offers. Unfortunately, these attributes, (as well as their weights), have not been selected through a rigorous assessment of social preferences. Thus, although the attributes might well be considered important, and a systematic and transparent system of weighted social preferences might help frame a more structured dialogue between stakeholders and health authorities and payers, much work is needed before such a system can be evolved, which can simultaneously deal with the major issues of data quality (lack of strong or convincing clinical evidence), opportunity cost (high price of the product, resulting in high incremental cost-effectiveness ratio) and equity of access.

It has taken many years to develop a measure of health benefit (the QALY) which is fit for purpose in many different decision contexts – and it still has many acknowledged imperfections. To move to an MCDA approach, a measure of a similar reliability must be developed for each of the new dimensions of value introduced into the analysis. Because new frameworks and decision models do not offer an immediate solution to the above mentioned problems, we suggest a different approach to improving the use of cost-effectiveness evaluation in this field. Instead of creating new instruments, we suggest that efforts should be directed at optimising the input to the cost-effectiveness evaluation, in order to identify the applications of the products which are most likely to be cost-effective, for example by limiting the use of these products to those patients that show maximum benefit and in such way (containing stringent starting and stopping rules) that the associated cost-effectiveness is still within reasonable limits or at least drops substantially.

> Based on the preceding analysis, we aim to provide guidance on coverage decisions for (orphan) medicinal products that show limitations of the clinical data at the time of launch, by developing a Simple Transparent Algorithmic Multidisciplinary Procedure (STAMP) for the purpose of optimising the input for the standard set of decision criteria (effectiveness, cost-effectiveness and budget impact).

3.2 Simple Transparent Algorithmic Multidisciplinary Procedure (STAMP)

General principles

Based on the fact that strong or convincing clinical evidence of the orphan medicinal product may be lacking at the time of marketing authorisation, combined with a high price and no current assessment framework that solves the problem of high ICERs, we promote the following structure for coverage decisions:

> ➢ an initial assessment based on the standard assessment criteria set 'effectiveness, cost-effectiveness and budget impact' for creating a reference point with respect to each criterion,
> ➢ a temporary coverage period in which it can be verified if the clinical evidence (including both effectiveness and quality of life) has improved or can be confirmed, and in which additional economic data can be gathered, and
> ➢ A final assessment based on the standard assessment criteria set containing the optimised input for evaluating effectiveness, cost-effectiveness and budget impact.

With respect to this structure, we identify five areas which together constitute the total 'simple transparent algorithmic multidisciplinary procedure' (STAMP) for coverage of orphan medicinal products. The areas are listed below and illustrated in Figure 3.

- Area 1: timing of the procedure
- Area 2: professional appraisal committee
- Area 3: stakeholder review and advice
- Area 4: conditional coverage instruments
- Area 5: (conditional) coverage decision-making by health authorities

The areas are discussed in section 3.4 'implementation'. In section 3.5 overall organisational aspects of the STAMP will be discussed.

In general, we consider the presented procedure 'STAMP' applicable to all orphan medicinal products, though most likely exceptions will exist. In section 3.4.2 we discuss exceptions to STAMP. Furthermore, the 'STAMP' is also considered applicable to non-orphan medicinal products that show limitations in the clinical data at the time of launch.

3.3 Illustration STAMP

Figure 3: Overview of the algorithmic approach in STAMP.
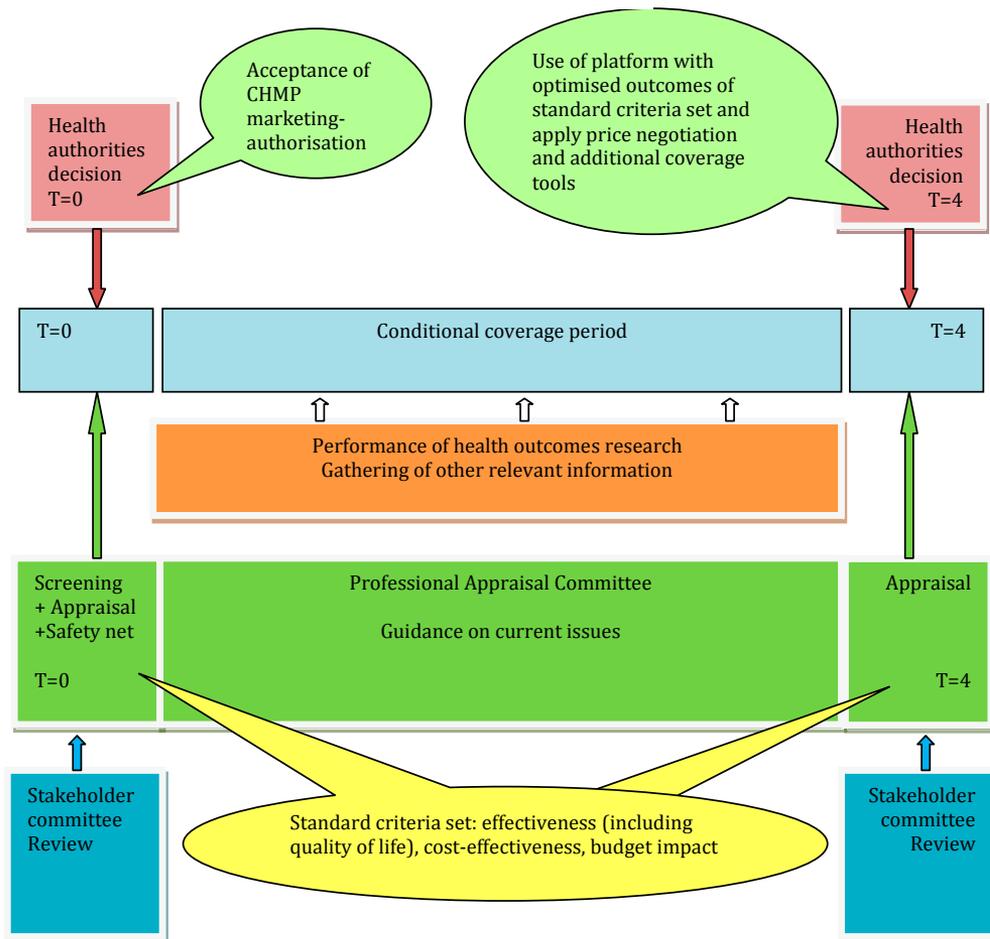
3.4 Implementation STAMP

3.4.1 Area 1: timing

Based on the fact that the European Committee has pushed the development and marketing of orphan medicinal products in the European Union, it can be stated that orphan medicinal products, as with all new products approved by EMA, should be made available to patients as soon as possible after marketing-authorisation (registration procedure) due to the fact that in most orphan diseases no alternative therapy is available. Therefore, the assessment for coverage of orphan medicinal products and decisions by health authorities need to be done soon after marketing-authorisation has been granted. This assessment, directly after marketing-authorisation, is indicated as T=0.

The Professional Appraisal Committee performs an initial screening (to see whether there is sufficient data available to carry out a meaningful evaluation) and subsequently an assessment at T=0. The screening and assessment procedures and outcomes are explained in section 3.4.2 and 3.4.5. Based on the outcomes the Professional Appraisal Committee will recommend health authorities that the product is provided definite coverage, temporary coverage or rejection from coverage.

During the temporary coverage, additional research needs to be done for the purpose of improving or confirming the clinical evidence (including effectiveness and quality of life), at the same time reducing the confidence interval in ICER and reducing the ICER.

A post-authorisation period of maximum 4 years is expected to be sufficient to generate additional data. The design of the health outcomes study should warrant a conclusion at T=4. After the temporary coverage period, a second assessment is done. The moment of this assessment is indicated as T=4. The assessment results in a final recommendation to health authorities on coverage. At the time a decision on coverage is taken by the health authorities, the temporary coverage is stopped.

3.4.2 Area 2: Professional Appraisal Committee

We suggest that a committee of multidisciplinary experts is established. It is recommended that such a committee operates as an independent group, not linked to specific clinical investigators, health insurers, health authorities or pharmaceutical industry. The responsibility of appointing such a committee should lie with the national branch organisation of clinicians, but we suggest that it should be supported by a government related body, for example, NICE in the UK or the CVZ in the Netherlands.
We suggest that one committee is installed for operating on all candidate medicinal products for this STAMP procedure. However, the composition of this committee may vary over time.

The composition of this external committee is suggested to be the following:
- Clinical expert in rare diseases but not in the specific orphan disease under review in order to avoid conflict of interest
- Economic expert
- Quality of life expert
- Expert on patient related aspects
- Epidemiological/methodological expert
- Ethical professional

Depending on the subject the committee may vary in type and number of members. The members of this committee should be specialists with extensive experience in their area, but should not have been involved in the development process of the (orphan) medicinal product or involved in the treatment of these patients. This provides a multidisciplinary, specialised but objective view on all available data and, most of all, an objective assessment of the standard assessment criteria and advice on other related tasks.

Information
The Professional Appraisal Committee will be provided with information on effectiveness, on the cost-effectiveness model (T-0 model) and on other relevant data by the manufacturer of the specific orphan medicinal product. Additional data or analyses can be requested from the manufacturer. Also, we emphasise that the members of the committee make every effort to obtain, in a systematic way, already available information from both national and/or European expertise centres if ikjappropriate (clinical and patient information). We suggest that the committee liaise with the European Network for Health Technology Assessment (EUnetHTA), the European Union Committee of Experts on Rare Diseases (EUCERD), and other existing national or European expert groups or HTA bodies that work on the methodology of evaluating orphan medicinal products, in order to align with current insights on the orphan medicinal product and disease area from a clinical and health economic perspective.[9][34][35][36]

Tasks
In summary we suggest that this committee should have the following tasks:

o Screening
First of all, the screening should be to see if there is sufficient data to do any evaluation. This situation should not arise for products licensed by EMA.
We recommend that the committee forms an opinion on the medical need the new orphan medicinal product fulfils. This is an essential step to protect the temporary coverage instrument from abuse by introduction of extremely expensive medicinal products which are not expected to fulfil a medical need or which are not expected to add value to the specific disease. For example, a new orphan medicinal product may be introduced as the industrial version of a pharmacy preparation at a very high price, while the original pharmacy preparation was already in common use for decades at a low price.

We suggest the committee aligns the outcome of the initial screening with the national government related body. If the product is judged not to fulfil a medical need, no STAMP will be started.

Another safety net for protecting the temporary coverage instrument from abuse by introduction of extremely expensive medicinal products, is the requirement of 'minimal evidence' at the T=0 assessment. This safety net is discussed in section 3.4.5 'coverage decision'.

○ At T=0, assessment using the traditional assessment criteria set for coverage (effectiveness, cost-effectiveness and budget impact).

If the orphan medicinal product fulfils a medical need, the committee starts the assessment procedure. It is suggested to apply the same standard assessment criteria set for coverage decisions, consisting of effectiveness, cost-effectiveness and budget impact, that is used for so-called 'mainstream medicinal products', i.e. medicinal products that are not hampered by limitations in the clinical data at the time of launch. This approach of applying the same standard assessment criteria set to all medicinal products, regardless of limitations in the clinical data at the time of launch, corresponds to EMAs' approach with respect to the marketing-authorisation procedure. All applications for marketing-authorisation, whether the products have an orphan designation or not, go through the same EMA committee and through the same assessment procedure – the Committee for Medicinal Products for Human Use (CHMP). The EMA states in the Orphan Regulation that no special assessment for marketing authorisation of orphan medicinal products exists and that such products should be subject to the same assessment as other medicinal products to guarantee that there will be no 'second rate' assessment. In applying the standardised assessment methodology and procedure for marketing-authorisation, the EMA has shown it is able to deal with product assessment situations in which the standard requirements cannot be met due to limitations of the clinical data at the time of launch. In their 'Guideline on Clinical Trials in small populations', it is concluded that there are no special methods for designing, carrying out or analysing clinical trials in small populations. Approaches are indicated for increasing the efficiency of clinical trials. It is recommended that the need for statistical efficiency should be weighed against the need for clinically relevant/interpretable results; the latter being the most important. In situations where obtaining controlled evidence on the efficacy and safety of a new treatment is not possible, the regulatory assessment is allowed to accept different approaches as long as it is ensured that the patients' interests are protected. In such situations, treatment conditions and data collection should be standardised and data should be of high quality and adhere to GCP standards.

We suggest that the Professional Appraisal Committee applies a similar approach for the assessment of orphan medicinal products in the STAMP procedure, i.e. application of standard assessment criteria combined with the necessary flexibility in the evaluation of the available data.

Whatever the type of orphan medicinal product, it is expected that the outcomes of the conventional analysis will be driven by the robustness of the clinical evidence and the actual

price of the product. Although the outcomes of these standard criteria might not show sufficient evidence for effectiveness and might not meet the national threshold for cost-effectiveness, at least outcomes are available that provide a reference versus standard therapy or best supportive care. Once orphan medicinal products have passed the screening for entering the STAMP, we suggest that based on the assessment outcomes and positive advice of the professional appraisal committee at T=0, health authorities accept the initial levels of clinical evidence and cost-effectiveness outcomes as interim outcomes that need further confirmation in the conditional coverage period.

The possible model for provision of conditional coverage is also found in the registration procedure of the EMA. Based on limitations of the clinical data at the time of launch, the EMA can grant conditional marketing-authorisation (MA), requiring yearly applications for continued MA by presenting new data requested by the CHMP. When sufficient data for a 'normal MA' are obtained the company gets 'full approval'. This is different from 'Marketing authorisation under exceptional circumstances', meaning that no further data can ever be provided due to the rarity of the condition and that it would not be ethically acceptable to withhold treatment from patients. For the latter type of MA, the CHMP performs yearly evaluations of the benefit/risk of the products for five years, following which it is decided whether MA can be continued. Follow-up safety studies/registries can be applied to all products and MA procedures when appropriate.

Based on the T=0 assessment outcomes for effectiveness and cost-effectiveness, a 'cost-effective price' (CEP) for the orphan medicinal product will be calculated. This is the price at which the maximum acceptable ICER threshold in the specific country is just met. It should be calculated using the standard evaluation methods for orphan medicinal products of the country concerned, e.g. the methods of the new NICE Highly Specialised Technologies Committee in England. The CEP at T=0 will be seen as a reference price. However, in this STAMP procedure it is proposed that orphan medicinal products are provided a temporary coverage period in which further clinical, quality of life and economic data are gathered for optimising the input to the standard assessment. In this temporary coverage period, the manufacturer gets the original price in exchange for the performance of additional gathering of data through a health outcomes program. Therefore the threshold price at T=0 is not imposed, but used as a reference price. Later on, the CEP at T=0 needs to be compared to the CEP calculated at T=4. The information about the threshold price at T=0 and T=4, as well as the outcomes of the standard assessment are used as inputs to the final coverage decision.

We illustrate the possible outcomes of the T=0 assessment in Table 4 in section 3.4.5 (area 5, coverage decision).

o <u>Definition of relevant gains for clinical evaluation and quality of life evaluation.</u>

Definition of a relevant gain means definition of what is regarded as a relevant improvement that should be achieved by the orphan medicinal product in the clinical area (practice) and/or

in the health-related quality of life area. Also cut-off scores for definition of a 'responder' are regarded as indicators for relevant gain. A relevant gain should be defined by the professional appraisal committee after consultation with clinicians and patients. It should be based on the change in health status and health-related quality of life which is needed to achieve cost-effectiveness at the reference price, allowing for the special considerations applied to orphan diseases. The threshold gain necessary to achieve cost-effectiveness (as defined for orphan medicinal products) at the manufacturers' price is the relevant gain to be considered. Together with the CEP, this can show the extent to which any post-launch research programme must change the outcomes, for reimbursement to be supportable. The more improved outcomes are at T=4 the less the manufacturer will have to reduce the price.

The defined relevant gains serve as thresholds for therapeutic value claims. Relevant gains also facilitate the identification of subgroups of patients that achieve this threshold for clinical added value or added value from a patient centred quality of life perspective. Furthermore they facilitate the development of starting and stopping criteria. Those patients that meet the defined relevant gains are those patients that benefit most from the treatment. If response to treatment does not reach the defined relevant gains, then this can be used as an indicator to stop treatment. Furthermore, these relevant gains will allow the establishment of a treatment protocol and the protocol for the health outcomes research during the conditional coverage period.

o   Advice on which patients to treat and the treatment protocol.

Based on the conclusions of the T=0 assessment, and definition of relevant gains, the professional appraisal committee should advise on the (sub) group(s) of patients to treat with the orphan medicinal product. For the identified group(s), advice on a treatment protocol would be provided, including, start- and stopping criteria.

o   Design of the health outcomes/registry program.

The professional appraisal committee provides suggestions for setting up the health outcomes/registry program. This program is executed within 4 years. The data would cover clinical effectiveness, quality of life and costs and other economic data. The collected data serve to optimise the input into the standard assessment at the point of re-evaluation. The health outcomes research may take the form of an observational study, a registry or a randomised phase 4 study. These new studies will address research questions and parameters that have already been part of the registration trial program or, if new research questions have arisen from the assessment process, new research questions might be added. Such health outcomes programs may cover countries other than those in the original registration studies. The additional data would provide further insights into the effects of the medicinal products in daily practice. At the end of the evaluation period, these data, together with any other newly available relevant data would form the input for the standard assessment at T=4.

The time period for this health outcomes research should be kept strictly to 4 years, and the assessment conducted with the best available data at that point.

Finally, it is important to have agreement beforehand, from all relevant stakeholders, on the design of the additional research program, and to take account of their views on the above issues. The criteria for the final decision at T=4 should also be defined in advance, based on analysis of various scenarios for the clinical and economic outcomes. The terms on which coverage might be stopped after four years should be defined, for example only continuation of treatment in responding patients.

Further details of the health outcomes program will be described under area 4 in section 3.4.4.

o   T=0 Consultation of the stakeholder committee.

The draft assessment and the outcomes of the other tasks performed by the professional appraisal committee will be made available for review by the stakeholder committee. This input is considered into the final T=0 report. The role of the stakeholder committee will be described under area 3 in section 3.4.3.

o   At T=0 advice on coverage to the health authorities.

Based on the outcomes of the standard assessment criteria at T=0, the professional appraisal committee would provide guidance for temporary coverage to health authorities at T=0. In addition, the committee would provide guidelines for the use of the medicinal product, for instance specific protocols for subgroups and start and stopping rules. This guidance would be published in a report for use by the health authorities and for external use.
The outcomes of the assessment of the standard criteria set and the threshold price, together with the definitions for relevant gains, treatment protocol, advice on guidelines and the design for the health outcomes program would serve as a plan for temporary coverage. This plan is described in more detail under area 4.

o   At T=4 assessment of the standard assessment criteria set for coverage (effectiveness, cost-effectiveness and budget impact).

After a period of 4 years (T=4) years, the professional appraisal committee will assess the health outcomes data together with other available data, by making use of the standard assessment criteria set (effectiveness, cost-effectiveness and budget impact). For this purpose all stakeholders should be asked to submit information, but the professional appraisal committee should decide what is admissible.

The professional appraisal committee will be provided with effectiveness data and an executable cost-effectiveness model (and data for processing purposes) by the manufacturer of the specific orphan medicinal product. The manufacturer is also responsible for providing

all the data from the health outcomes program to the expert committee. Additional data or analyses can be requested of the manufacturer.

All data gathered in the temporary coverage period is analysed. The outcomes of the T=4 assessment functions as a basis for final decisions on coverage.

The draft assessment performed by the committee will be made available for review by all relevant stakeholders. This input is considered in the final T=4 assessment. The committee would provide guidance on final coverage to health authorities. This guidance would be published in a report for use by the health authorities and for external use.

We illustrated the possible outcomes of the T=4 assessment in Table 5 in section 3.4.5 (area 5, coverage decision).

3.4.3 Area 3: Stakeholder review and advice

The draft output of the professional appraisal committee is reviewed and commented upon by stakeholders. This group of stakeholders would include the manufacturer, health authorities, specialised physicians and patient organisations. In a hearing the group of stakeholders expresses their comments and advises on the draft report of the professional appraisal committee. The external committee processes the comments and takes final decisions on the outline of the advice to the health authorities.

It is important that this review committee of stakeholders includes specialised physicians, patient organisations and manufacturer. While it is likely that these committee members may be less objective due to their strong interest in the product be covered, their expertise on the specific orphan medicinal product would be of help in providing detailed feedback to the professional appraisal committee. Furthermore, given their expertise they would have significant advantage in judging the achievement and function of the health outcomes program during the conditional coverage period. The exchanging of information and commenting would be carried out in a hearing in which both committees (professional appraisal and stakeholder committee) would attend at T=0 and T=4.

3.4.4 Area 4: conditional coverage

Definition
The definition of conditional coverage is temporary coverage with the agreement that health outcomes research will be conducted, the results of which will be used to judge whether permanent coverage should be granted. Research objectives would focus on generating additional input for the assessment of effectiveness and cost-effectiveness. In this way, the final decision on coverage by heath authorities is postponed until more information is available. This reduces the commitment by health authorities to spend health care budget on high priced orphan medicinal products, without having more certainty about the effectiveness and cost-effectiveness outcomes.
Furthermore it is a measure to prevent false-positive and false-negative coverage decisions. A false

positive decision is acceptance for coverage while at a later stage it will become clear that the orphan medicinal product does not show convincing clinical effectiveness and or cost-effectiveness; a false negative decision is a rejection of coverage while at a later stage it will become clear that the orphan medicinal product does show convincing effectiveness and or cost-effectiveness in the whole patient population or in subgroups.

Role and responsibility

It will be the responsibility of the professional appraisal committee to provide a design for the health outcomes program. The participation of physicians and patients is mandatory for temporary coverage. Specialists, patient organisations and manufacturer are responsible to provide their support to the execution of this program. Alignment is needed on the design, the execution and the access to information. The specialised centres in this specific orphan disease must be involved in the health outcomes research. Preferably, the health outcomes research is based on a European program executed in not only the national centres but also other foreign centres, to increase the sample size. We recommend that an independent research organisation carries out the organisation and gathers the data.

Research organisation

The execution of a health outcomes program involves several stakeholders. Currently, we have seen various ways of organising such health outcomes programs. Some professional societies of specialists install and execute health outcomes programs for several diseases relating to their profession. They build databases in order to evaluate their own professional practice and guidelines. In return for grants, other stakeholders may ask for access to information. Other professional societies perform one particular health outcomes program on an ad hoc basis and seek funding at a national governmental fund organisation. In some cases, the manufacturer takes the initiative to set up a health outcomes program, often as a part of a larger European or global registry that is initiated by headquarters. When the manufacturer is the owner of the health outcomes program and its information, the access to the information might be hampered.

Therefore, we recommend that an independent research organisation is appointed to execute the health outcomes program. This organisation may also perform data analysis and evaluation. It is also recommended that the database is owned by the research organisation that executes the health outcomes program. The research organisation may provide access to information on request to relevant stakeholders.

We suggest that a direct link between the funders and the research organisation is avoided. Therefore we recommend that a fund is raised from, preferably, various stakeholders, from which the health outcomes program is financed.

Design and study protocol

The design and study protocol for the health outcomes research should be defined by the professional appraisal committee. This will involve definition of the patient population, inclusion and exclusion criteria, selection of appropriate clinical outcome measures, including the selection of appropriate (patient related) Quality of Life scales and definition of relevant gains of these parameters, comparative treatment, size of the cohort, assumptions and methodological considerations (subgroups of patients and start- and stopping criteria). The health outcomes

research is based on the clinical parameters and quality of life parameters already used in the pivotal clinical studies and the economic parameters used in the cost-effectiveness evaluation at T=0. If new research questions have arisen from the assessment process, new research questions might be added.

The gathering of data will be done preferably prospectively, although information about control patients might be gathered retrospectively. The data might be gathered in targeted longitudinal observational databases, or patient registries, which reflect the current treatment patterns without influencing the treatments or interventions. However, randomised phase IV studies might be an option as well. It is important to discuss any methodological concerns, if a RCT is not feasible, and possible solutions to be defined on beforehand instead of during the post-hoc phase of the study. If RCT is considered the only option, the set-up of an international study may be considered in order to increase sample size. Data may also come from other sources like international clinical trials, safety programs and patient-information.

Platform of information

The outcomes of the health outcomes program provide a platform of information:

- Information on subgroups for treatment

- Start and stopping criteria

- For guideline optimisation

- For optimisation of the input for the standard assessment criteria

- For defining the threshold price at T=4.

- Subgroups of patients for further future application of additional coverage instruments like stratified medicine, value based pricing (VBP), patient access scheme, pay-for-performance and prescription restrictions.

At the end of the 4 years, the health outcomes program will bring together all the data gathered in the temporary coverage time period (including any available from other studies) and will constitute the input for the T=4 appraisal.

Funding

Observations of several health outcomes programs in the Netherlands showed that these programs were executed with either public or private funding or a mixture of both. In most cases of public funding, the funding does not cover all expenses of the program. In such cases support was requested from manufacturers. When the health outcomes program was funded completely by the manufacturer, the manufacturer was also owner of the program and subsequently of the information coming from it, hampering access by other stakeholders. Other examples of private funding by the manufacturer consist of providing grants to professional societies which hold health outcomes programs themselves, often on a large scale, covering more diseases. Also here, access to information by other stakeholders is limited.

Therefore, funding of the health outcomes program for the purpose of conditional coverage is the responsibility of the manufacturer in return for temporary coverage by health authorities. Besides

the funding of the health outcomes program, we also mention here the funding of the work of professional appraisal committee, which is regarded as the responsibility of the manufacturer as well. Corresponding to the EMA model and also local registration authority models, funding of the assessment activities is paid for by a system of fees for entry by the manufacturer (see also section 3.5). We suggest that the control of the budget and activities should lie with the professional appraisal committee. In case of shortfall, back-up public funding might be provided by health authorities.

3.4.5 Area 5: coverage decision-making by health authorities

We distinguish two time points for coverage decisions by the health authorities: T=0 and T=4.

Coverage decision at T=0
At T=0 the health authorities receive the report of the professional appraisal committee with their advice. The possible outcomes of the assessment at T=0 by the professional appraisal committee are illustrated in Table 4.

Table 4: Overview of possible outcomes at T=0, using the standard criteria effectiveness and cost-effectiveness.

Cost-effectiveness outcomes / Effectiveness outcomes	Outcome < ICER threshold	Outcome ≥ ICER threshold
Sufficiently proven	+ +	+ -
Not sufficiently proven	- +	- -

In case a product meets both effectiveness and cost-effectiveness levels (option +, + in Table 4), we suggest that the product is accepted for permanent coverage and does not enter the conditional coverage procedure.

For the remaining three cases, still the outcomes show uncertainty in effectiveness and/or cost-effectiveness. To prevent abuse of the temporary coverage opportunity, we recommend to apply a safety net, despite the knowledge that these products have passed already the initial screening for medical need.

Safety net
This safety net consists of testing the outcomes of cost-effectiveness at T=0 for 'minimal evidence' before these products enter into a conditional coverage procedure. One recommendation to test the minimal evidence is through the construction of an acceptability curve. We assume that all uncertainty related to limited clinical evidence is reflected in the acceptability curve. Therefore, the acceptability curve may be used as a criterion for entering the temporary coverage. E.g. a

requirement for entering the temporary coverage may be: the probability of ≤ 10% that the ICER is higher than e.g. euro x,000. The limited clinical evidence may lead to large confidence interval for the ICER and therefore this approach may be justified. In this approach we assume all uncertainty is included in the distributions. We also assume that relevance of outcome is reflected in uncertainty in utilities.

Figure 4: An example of a cost-effectiveness acceptability curve.
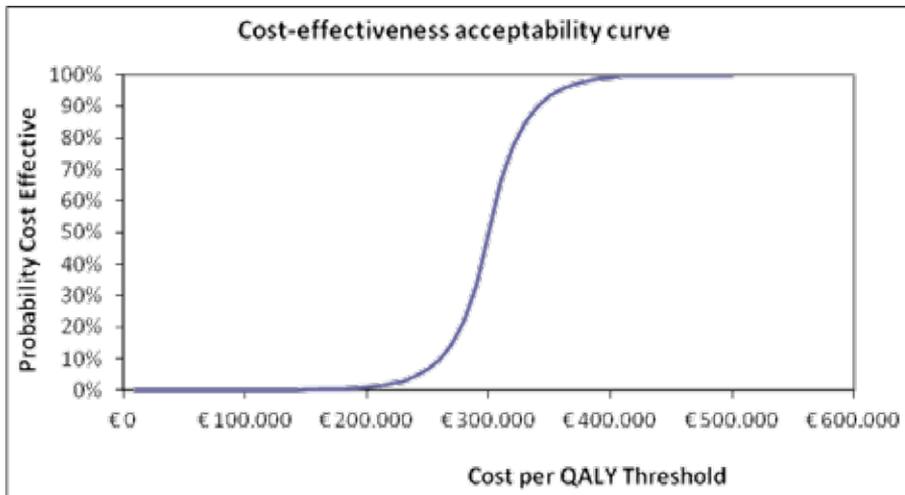
Another recommendation to test the minimal evidence is using extensive one-way and two-way sensitivity analyses on the key criteria (limited clinical evidence and high price), which drive the ICER. Limited clinical evidence is translated in T=0 model into transition probabilities and utilities. These sensitivity analyses may show that even extreme sensitivity analyses may never lead to acceptable ICER, even if we assume a higher ICER threshold for orphan medicinal products. Either an extremely unrealistic efficacy gain is required, or an unrealistically low price for the orphan medicinal product is required. If at realistic assumptions for efficacy, the price is the only issue, and e.g. 30% lower price will make the product cost-effective, than it is important to have commitment upfront from the manufacturer that indeed they will be willing to lower the price at T=4. If not, then there is no reason to start STAMP. In other words, if there is a negligible chance that the product will meet reimbursement criteria despite much additional data being gathered, then it should not enter STAMP.

Therefore, for the remaining three cases (option -, + and -, + and -,- in Table 4), we suggest uptake of the product into a conditional coverage procedure after they have passed the safety net as described above. Even the case (option -,- in Table 4) may lead at T=4 to a sufficiently proven evidence base and an acceptable ICER. We suggest that based on the assessment outcomes and positive advice of the professional appraisal committee at T=0, health authorities accept the initial levels of clinical

evidence and cost-effectiveness outcomes as interim outcomes that need further confirmation in the conditional coverage period. Apart from the advice of the committee, health authorities may apply their own policy criteria.

Coverage decision at T=4

At T=4 the health authorities receive the report of the external committee containing the outcomes of the standard assessment criteria based upon all input information gathered through the health outcomes program. The evaluation of the information starts with the basis 'effectiveness', and more specifically with the achievement of relevant gains. We can distinguish several scenarios with respect to effectiveness after 4 years (T=4):

- No clinical evidence
- Still insufficient convincing clinical evidence not meeting relevant gain definitions
- Convincing clinical evidence meeting relevant gain definitions

The outcomes with respect to effectiveness are, in particular, leading for defining which patients benefit most, for definition of the treatment protocol, for start- and stopping criteria, cost-effectiveness and budget impact. Each of the three scenarios is considered in more detail below.

T=4 assessment - no clinical evidence

If the initial, insufficiently-convincing clinical effects are not reproduced and if the defined relevant gains are not met, the product will no longer be regarded as having a sufficient therapeutic value, therefore not requiring any further coverage. As mentioned the consequences of stopping coverage after four years should have been defined at T=0 and are now executed. The cost-effectiveness outcome is not relevant anymore in this case. As there is no actual therapeutic value at T=4, any newly calculated ICER will still have a large confidence interval.

T=4 assessment - still insufficiently-convincing clinical evidence not meeting relevant gain definitions

The T=0 assessment with the definition of relevant gains might have provided the platform for identifying subgroups of patients that do show some improvement of the clinical effect compared to the level of effect at marketing-authorisation, however at T=4, the clinical and quality of life outcomes do not meet the relevant gains for effectiveness.

When the relevant gains are not met, the ICER will remain higher than the ICER threshold. That means that not sufficient reduction in ICER could have been created by optimising the use of the orphan medicinal product. Depending on the size of the ICER, and the difference in relation to the highest acceptable ICER in a specific country, coverage could only be afforded if the price of the orphan medicinal product is lowered to the threshold price. The new price might be a reference price for this type of orphan medicinal product. At this stage, it is recommended that manufacturer and health authorities discuss upon the final coverage possibilities for the orphan medicinal product. Based on the effort that has been put into optimisation of the input for the standard assessment criteria and based on the outcomes of the standard assessment criteria at T=0 and T=4, both manufacturer and health authorities have a transparent platform for discussing price and application of additional policy tools including price-arrangements like price-volume arrangements.

Because of forthcoming data of the medicinal product in real practice for this subpopulation, the value assessment may not stop at T=4, but may be repeated at T=6 based on newly achieved knowledge about appropriate biomarkers for measuring effects of the medicinal product. Especially for the identification of the subpopulation, biomarkers may become critical for further coverage. Especially in orphan diseases it is of importance to gather life-long data for the purpose of understanding, for example, disease progression. However, continuation of the health outcomes program does not imply continuation of conditional coverage. At T=4, conditional coverage is discontinued, regardless whether enough data has been gathered.

T=4 assessment - convincing clinical evidence meeting relevant gain definitions
The orphan medicinal product has confirmed the clinical evidence available at the time of marketing-authorisation and has met the defined relevant gains, at least for a subgroup of the population. For coverage decisions it is important that the defined relevant gain has been achieved. Meeting the defined relevant gain(s) means that therapeutic added value of the orphan medicinal product is confirmed.
The ICER will be calculated based on the actual clinical and quality of life evidence. Based on the optimisation of input on effectiveness and cost-effectiveness, it is expected that the new calculated ICER is less than the ICER at T=0. At the same time, the calculated threshold price of the orphan medicinal product may have become higher than the threshold price at T=0. The threshold price may serve as new threshold price for this type of orphan medicinal product and indication.

The possible outcomes of the assessment at T=4 by the professional appraisal committee are illustrated in Table 5.

In case the relevant gains are not met and the ICER remains higher than the ICER threshold (option -,- in Table 5). This means that although the use of the orphan medicinal product might have been optimised, more certainty is achieved that the orphan medicinal product does not produce the necessary gain. We suggest that the product is excluded from coverage. In case the relevant gains are not met, but the ICER is within acceptable ranges (option -,+ in Table 5), we also suggest to exclude the product from coverage. The latter option seems more or less a theoretical option.

If the defined relevant gains are met and the ICER is within acceptable ranges (option +,+ in Table 5), valid in a specific country, the orphan medicinal product may be accepted for permanent coverage. The use of the maximum upper limit is usually based on the inclusion of social values, burden of diseases, and specific ethical issues for orphan drugs, as described in previous section on the assessments of orphan drugs by SMC and NICE. In case the medicinal product meets all requirements, its threshold price most likely corresponds to the actual price of the orphan medicinal product.

If the defined relevant gains are met and the ICER is still above acceptable ranges (option +,- in Table 5), it is likely that the threshold price differs from the actual price of the orphan medicinal product. In this case we suggest that the manufacturer and health authorities discuss upon the application of policy tools including price- and/or volume arrangements and value based pricing.
Provided that sufficient effort is put in the health outcomes program and optimisation of subgroup determination and start and stopping criteria, the comparison of the outcomes for the ICER's and

threshold prices at T=0 versus T=4, illustrates the best achievable cost-effectiveness outcomes for that specific orphan medicinal product. This provides a transparent platform for manufacturer and health authorities for discussing upon price and upon application of additional policy tools, if necessary, during the decision-making process for permanent coverage.

Table 5: Overview of possible outcomes at T=4, on the standard assessment criteria effectiveness and cost-effectiveness.

Cost-effectiveness outcomes Effectiveness outcomes measured by relevant gain reached	Outcome ≤ ICER threshold (+)	Outcome > ICER threshold (-)
Defined relevant gain reached	+ +	+ -
Defined relevant gain not reached	- +	- -

Budget impact

Budget impact is one of the main standard criteria for coverage assessment. Health authorities decide whether the budget impact can be afforded within the health care budget. The forecast budget of the subgroup of patients, that met the effectiveness requirements, i.e. achievement of relevant gains, can be used for further fine-tuning the ICER threshold price. As such, the budget impact is used in coverage agreements, and can be managed by price-volume agreements, maximisation of volume/budget/costs per patient, number of treatments per patient or number of healthcare providers. Furthermore, a budget impact assessment should take account of the actual treated patient population and any likely off-label use.

Coverage agreements

The application of defined relevant gains in the assessment of effectiveness (clinical and quality of life perspectives), and for determining subgroups of patients that benefit most and start- and stopping criteria, provides insight into the possible options for a more efficient use of expensive medicinal products. Subsequently, through the health outcomes program, the cost-effectiveness outcomes can be optimised. The advantage of the STAMP approach is the transparency in the procedure and the maximum effort that is put into processing the four main issues, mentioned earlier, 'lack of strong convincing clinical evidence, high product price, uncertainty in the confidence interval of the ICER/high ICER' and 'lack of alternative treatments'. This approach provides the necessary platform for final coverage decisions and possible application of instruments to modify utilisation, like stratified medicine, value based pricing, prescription restriction and risk sharing agreements.

How the outcomes of the conditional coverage period will be validated and implemented into coverage agreements, depends largely on the local (national) policy and health budgets possibilities.

3.5 Organisation

Other organisational aspects of the STAMP approach, that have not been discussed above, are discussed here. They include:
- Scope, orphan medicinal products
- Central procedure
- Role of health authorities
- Role of the stakeholder committee
- Funding

Orphan medicinal products
We have recommended in prior sections, that the professional appraisal committee determines whether the orphan medicinal product under assessment fulfills a medical need. We suggest this is checked with the government related body. If, the product does not fulfil a medical need, it will fail the T=0 assessment and therefore no assessment will be performed as well as no coverage procedure followed. All other orphan medicinal products will be enrolled in this STAMP procedure and provided preliminary coverage.

Central procedure
It is recommended to have one central, national assessment procedure of orphan medicinal products. However, opposite to the connotation of 'central', the procedure is not carried out by a governmental body, but as described earlier, by an independent professional appraisal committee. We recommend that the national branch organisation of physicians/specialists selects the members and installs the professional appraisal committee. The branch organisation of physicians/specialists and the professional appraisal committee develop the STAMP in more detail. The outcomes of the STAMP procedure need to be included in the policy assessment procedure of health authorities. The committee is responsible for the assessment of orphan medicinal products at T=0 and T=4. It is also this committee that is in control of paying the independent research organisation that executes the health outcomes program. Responsibilities for facilitating the process of optimising the input for the standard assessment criteria effectiveness and efficiency (cost-effectiveness) as well as the assessment of these criteria belongs to the professional appraisal committee, while health authorities are responsible for final policy decisions according to national policies. In this way transparency is kept for all stakeholders with respect to 'content' and 'policy'.

Health authorities
Health authorities are involved in the stakeholder committee. Furthermore this procedure anticipates at the coverage decision by health authorities, in our view preferably at T=0 and T=4, due to the proposed conditional coverage prior to a final coverage decision.

Group of stakeholders
This group is installed by the professional appraisal committee. The stakeholders are notified at the beginning of the assessment procedure and are kept informed during the assessment procedure. At two points in the assessment process the stakeholders would be formally involved and have the opportunity to comment upon the advice of the professional appraisal committee.

Funding

We suggest that manufacturers are responsible for payment of the STAMP procedure (including the health outcomes program) in exchange for temporary coverage at the original price of the orphan medicinal product. We suggest that for the assessment procedure a fixed amount per orphan medicinal product is paid, comparable to the principle that is applied in registration procedures. For the health outcomes program the amount paid may differ depending on the extensiveness of the program. We suggest that the payments are made to a fund that is kept by the body managing the professional appraisal committee, i.e. NICE in the UK. From this fund all activities of the professional appraisal committee and the research organisation that executes the health outcomes program are paid, as well as the accommodation for meetings and office facilities.

Table 6: Overview organisation, committees, stakeholders, health authorities and tasks.

Stakeholders	Appraisal T=0 & T=4	Review Appraisal T=0 & T=4	Health Outcomes Research			Decision conditional coverage	Decision final coverage	Funding STAMP Fee per procedure
			Research	Funding	Organisation			
Health authorities		X				X	X	
Professional Appraisal Committee -Clinician -Economist -QoL expert -Methodologist/Epidemiologist -Patient expert -Ethical expert	X		X		X			
Stakeholder Committee -Manufacturer -Health Authorities -Health Insurers -Specialists -Patient societies		X						
Health Outcomes Research Organisation		X	X					
Manufacturer		X		X				X

3.6 Additional issues for further considerations

Price of orphan drugs:

More transparency of costs of manufacturing (R & D costs) by manufacturers is more and more suggested. The risk sharing agreements may also require the submission of an accounting report

documenting the high production and development cost of the new orphan drug. This option becomes especially relevant, if even in case of clinical evidence, the threshold price based on the ICER remains below the break-even price for the manufacturer. In this situation, the payers may be willing to accept a higher drug price, even if the ICER exceeds now the upper limit.

Collaboration between EMA/FDA and national HTA bodies
The EMA opinion-making process for marketing-authorisation of orphan medicinal products applies the same standard criteria for assessment for all medicinal products, including orphan medicinal products. Within this standard assessment framework the EMA has a step by step approach. If for example RCT's are lacking, the EMA considers the next best level of evidence, and works with this evidence. However, prerequisite is that the quality of the data is sufficient and the way this data have been acquired is along the correct clinical practice standards. Based on the scarce but qualitatively qualifying data, the EMA gives opinion on marketing-authorisation of the orphan medicinal product, followed by decision by the European Commission. Health care authorities however, tend to apply the standard assessment criteria for coverage to orphan medicinal products without flexibility to approach the level of evidence on a step by step basis. This example shows that more harmonisation between EMA registration authorities and health care authorities is required in order to develop consistent policy rules for covering orphan medicinal products.

Lifelong follow-up
Apart from coverage decisions and health outcomes programs, which are set at T=4 (years), society has an interest in continuing research. Therefore, it is suggested to generate lifelong follow-up programs to gather information about the orphan disease itself and its treatments for better understanding the disease. In case of very expensive medicinal products, this effort may be required from physicians and patients, supported by health research organisations and HTA bodies. Many orphan diseases lack information about the progress of the disease and specific treatment aspects that make it difficult to interpret the effect size of a new therapy. As is done in the orthopedic area, where lifelong information about hip and knee replacement is gathered, it is appropriate to extend the health outcomes research. Subsidies from independent institutes and funds could provide the necessary budget for these follow-up programs.
In addition the forthcoming results of biomarkers may become available and may lead to adjustment of the treatment guidelines. Because of these continuing developments and forthcoming data of the orphan medicinal product in real practise, the value assessment may not stop at T=4, but may continue over time at predefined intervals. Only new RCT data may be the basis for extension of the patient population, but this additional population should follow the same proposed procedure from T=0 until T=4, and may be considered as a separate procedure.

3.7   Conclusions

In the case of orphan medicinal products that show limitations of the clinical data at the time of launch, alternative assessment frameworks and policy tools do not solve the basic problems of 'lack of strong or convincing clinical evidence', and 'high price of the medicinal product', resulting in a large confidence interval in the ICER.

We suggest that efforts should be directed at optimising the input for the existing standard assessment process.

We developed a <u>S</u>imple <u>T</u>ransparent <u>A</u>lgorithmic <u>M</u>ultidisciplinary <u>P</u>rocedure (STAMP) to provide guidance on how to optimise the input for the standard assessment of effectiveness, cost-effectiveness and budget impact', in (orphan) medicinal products that have limited clinical data available at the time of launch.
The STAMP provides a structure in which responsibilities of stakeholders are transparent,   with open access to research information, assessment information is available, and 'assessment of content' is separated from 'policy decisions'. The STAMP also integrates   the possibility to align with already existing networks and centres of expertise. We conclude by stating that this report used orphan medicinal products as an example, however the STAMP is applicable for all medicinal products that show limitations of the clinical data at the time of launch.

Declaration of interest

Coordinator expert committee:
M. (Mark) J.C. Nuijten, MD, PhD, MBA   Amsterdam,
                                                The Netherlands

No relationships/conditions/circumstances that present a potential conflict of interest.

Expert committee members:
Prof. J (John) Hutton, BPhil, Health Economist   University of York, York,
                                                  United Kingdom

No relationships/conditions/circumstances that present a potential conflict of interest.

Prof. F. (Frans) H. Rutten, PhD, Health Economist   Institute for Medical Technology Assessment
                                                     (iMTA), EUR Rotterdam,
                                                     The Netherlands

No relationships/conditions/circumstances that present a potential conflict of interest.

P. (Peter) M. van Hasselt, MD, PhD   University Medical Centre Utrecht, for
Pediatrician for metabolic diseases   metabolic diseases,
                                              The Netherlands

No relationships/conditions/circumstances that present a potential conflict of interest.

C. (Cor) Oosterwijk, PhD   Dutch Genetic Alliance (VSOP), Soest,
                                    The Netherlands

No relationships/conditions/circumstances that present a potential conflict of interest.

D. (Dominique) J. Dubois, MD, FFPM, FBCPM   Université Libre de Bruxelles (ULB),
Physician Specialist in Pharmaceutical Medicine   Belgium

No relationships/conditions/circumstances that present a potential conflict of interest.

Advisors
Mrs. H. (Hanka) K. Meutgeert   VKS (Patient – Association for metabolic
                                        disorders), Zwolle,
                                        The Netherlands

No relationships/conditions/circumstances that present a potential conflict of interest.

References

[1] Rabinovich,M.,Greenberg,D.,and Shemer,J. (2007).Threshold values for cost-effectiveness ratio and public funding of medical technologies. Harefuah 146, 453–458, 500.

[2] Nuijten MJ, Dubois DJ. Cost-utility analysis: current methodological issues and future perspectives. Front Pharmacol. 2011 Jun 8;2:29. doi: 10.3389/fphar.2011.00029. eCollection 2011.

[3] Zicht op zinnige en duurzame zorg. Zoetermeer: Raad voor de Volksgezondheid en Zorg, 2006.

[4] http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf).

[5] http://www.emea.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000029.jsp&mid=WC0b01a

[6] European Medicines Agency. EPAR Myozyme. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Scientific_Discussion/human/000636/WC500032128.pdf.

[7] College voor zorgverzekeringen. Diemen 2012. Report assessment health outcomes research alglucosidase alfa (Myozyme®) for Pompe disease. Diemen: CVZ; 2012. http://www.cvz.nl/binaries/live/cvzinternet/hst_content/nl/documenten/rapporten/2012/rpt211-ziekte-van-pompe.pdf.

[8] Dubois in M. Mesbach et al. (eds.). Statistical Methods for Quality of Life Studies, 307-308.

[9] EUnetHTA JA1 WP5 methodology guidelines Endpoints used for relative effectiveness assessment of pharmaceuticals HEALTH-RELATED QUALITY OF LIFE and UTILITY MEASURES. February 2013.

[10] College voor zorgverzekeringen. Diemen 2012. Report assessment health outcomes research agalsidase alfa (Replagal®) and agalsidase bèta (Fabrazyme®) for Fabry disease. Diemen: CVZ; 2012. http://www.cvz.nl/binaries/live/cvzinternet/hst_content/nl/documenten/rapporten/2012/rpt1211-ziekte-van-fabry.pdf.

[11] Goossens ME, Rutten-van Molken MP, Vlaeyen JW, van der Linden SM. The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. J Clin Epidemiol 53:688–695, 2000

[12] National Institute for Health and Care Excellence. Appraising Orphan Drugs. Last updated: 14 April 2008. www.nice.org.uk/aboutnice/whoweare/seniormanagementteam/meetings/2005/12july2005/appraising_orphan_drugs.jsp; www.nice.org.uk/niceMedia/pdf/smt/120705item4.pdf.

[13] Letter of Ministry of Health 29 January 2013. The Netherlands

[14] Letter of Ministry of Health 3 October 2013. The Netherlands.

[15] Scottish Medicines Consortium. SMC advice on alglucosidase alfa (Myozyme®) (352-07) SMC Drug ID: 352/07; Manufacturer: Genzyme Therapeutics Ltd; Indication: Pompe disease; Submission Type: Full submission; Status: Not Recommended; Date Advice Published: 12/03/2007.

[16] Hughes-Wilson W, Palma A, Schuurman A, Simoens S. Paying for the Orphan Drug System: break or bend? Is it time for a new evaluation system for payers in Europe to take account of new rare disease treatments? Orphanet J Rare Dis. 2012 Sep 26;7:74. doi: 10.1186/1750-1172-7-74.

[17] Denis, et al: Budget impact analysis of orphan drugs in Belgium: estimates from 2008 to 2013. J Med Eco 2010, 2: 295–301.

[18] Belgian Federal Centre for Healthcare, Belgian Health Care Knowledge Centre: Policies for Rare Diseases and Orphan Drugs. KCE Report 112C"2009, Also published as: Denis A, Mergaert L, Fostier C, Cleemput I, Simoens S. "Budget impact analysis of orphan drugs in Belgium: estimates from 2008 to 2013. J Med Econ 2010, 13(2):295–301.

[19] ALCIMED: Study on orphan drugs. 2004. Phase I, 2004 http://ec.europa.eu/ health/files/orphanmp/doc/pricestudy/final_final_report_part_1_web_en.pdf.

[20] Carina S, Tsveta M, Adam H: Estimating the budget impact of orphan medicines in Europe: 2010 – 2020. Orphanet J Rare Dis 2011, 6:62.

[21] Sussex J, Rollet P, Garau M et al. Multi-Criteria Decision Analysis to Value Orphan Medicines. Research Paper 13/03. OHE. May 2013.

[22] Nuijten MJ, Mittendorf T, Persson U. Practical issues in handling data input and uncertainty in a budget impact analysis. Eur J Health Econ. 2011 Jun;12(3):231-41. doi: 10.1007/s10198-010-0236-4. Epub 2010 Apr 3.

[23] Trusheim MR, Berndt ER, Douglas FL. http://www.ncbi.nlm.nih.gov/pubmed/17380152.

[24] Deverka PA et al. Economic Opportunities and Challenges for Pharmacogenomics; Annual Revue of Pharmacology and Toxicology 50: 423-37, 2010.

[25] Eugene F. Diekman, TomJ. de Koning, Nanda M. Verhoeven-Duif, MaroeskaM. Rovers, Peter M. van Hasselt. Survival and Psychomotor Development With Early Betaine Treatment in Patients With Severe Methylenetetrahydrofolate Reductase Deficiency. JAMA Neurol. doi:10.1001/jamaneurol.2013.4915.Published online December 9, 2013.

[26] Scott MG. When do new biomarkers makes economic sense? – Scandinavian Journal of clinical & Laboratory Investigation, 2010; 70.

[27] Shah KK. Severity of illness and priority setting in healthcare: a review of the literature. Health Policy. 2009 Dec;93 (2-3):77-84. doi: 10.1016/j.healthpol.2009.08.005. Epub 2009 Aug 27. Review.

[28] Garau M, Shah KK, Mason AR, Wang Q, Towse A, Drummond MF. Using QALYs in cancer: a review of the methodological limitations. Pharmacoeconomics. 2011 Aug;29(8):673-85. doi: 10.2165/11588250-000000000-00000.

[29] Sloan FA, valuing health care. Costs, benefits and effectiveness of pharmaceuticals and other medical technologies. Cambridge: Cambridge University Press, 1995.

[30] Rosenbaum PR. Discussing hidden bias in observational studies. Ann Intern Med 1991;115(11):901-5.

[31] New House JP, McCellan. Econometrics in outcomes research: the use of instrumental variables. Annu. Rev. Public Health. 1998. 19: 17-34.

[32] Bowden RJ, Turkington DA. 1984. Instrument Variables. Cambridge: Cambridge Univ. Press.

[33] Hine LK, Laird N, Hewitt P and Chalmers T. Meta-analytic evidence against prophylactic use of lidocaine in acute myocardial infarction. Archives of Internal Medicine 1989: 149: 310 1433-1438.

[34] Health Technology Authorities, EUnetHTA (http://www.eunethta.eu/). Mechanism of Coordinated Access(MoCA) (http://ec.europa.eu/enterprise/sectors/healthcare/competitiveness/process_on_corporate_responsibility/platform_ access/index_en.htm).

[35] Clinical Added Value of Orphan Medicinal Products – Information Flow CAVOMP-IF (http://www.eucerd.eu/?page_id=13). Joint Action European Medicines Agency (EMA).

[36] International researcher-team under the coordination of Professor Ralf-Dieter Hilgers of the Uniklinik RWTH Aachen, Germany. IDEAL: Integrated DEsign and AnaLysis of small population group trials. The research is funded by the 7.th Framework Programme of the European Union (FP7-HEALTH-2013-INNOVATION-1, No. 602 552). http://www.rwth-aachen.de/go/id/ehgl/?lidx=1.