University of Groningen

Health economic evaluation of gene replacement therapies

Aballéa, Samuel; Thokagevistk, Katia; Velikanova, Rimma; Simoens, Steven; Annemans, Lieven; Antonanzas, Fernando; Auquier, Pascal; François, Clément; Fricke, Frank-Ulrich; Malone, Daniel

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Health economic evaluation of gene replacement therapies: methodological issues and recommendations

Samuel Aballéa , Katia Thokagevistk , Rimma Velikanova , Steven Simoens , Lieven Annemans , Fernando Antonanzas , Pascal Auquier , Clément François , Frank-Ulrich Fricke , Daniel Malone , Aurélie Millier , Ulf Persson , Stavros Petrou , Omar Dabbous , Maarten Postma & Mondher Toumi

Published online: 11 Oct 2020.

Submit your article to this journal

Article views: 416

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

GUIDELINE

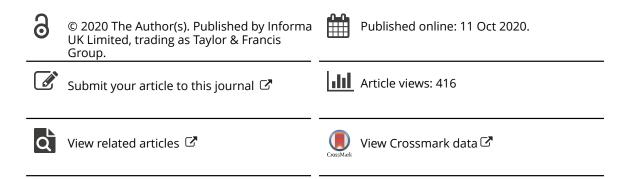OPEN ACCESS | Check for updates

Health economic evaluation of gene replacement therapies: methodological issues and recommendations

Samuel Aballéa[a], Katia Thokagevistk[b], Rimma Velikanova[c], Steven Simoens[d], Lieven Annemans[e], Fernando Antonanzas[f], Pascal Auquier[g], Clément François[b,h], Frank-Ulrich Fricke[i], Daniel Malone[j], Aurélie Millier[b], Ulf Persson[k], Stavros Petrou[l], Omar Dabbous[m], Maarten Postma [c] and Mondher Toumi [b,h]*

aCreativ-Ceutical, HEOR, Rotterdam, Netherlands; bCreativ-Ceutical, HEOR, Paris, France; cGroningen Research Institute of Pharmacy, PharmacoTherapy, Epidemiology & Economics, University of Groningen, Groningen, Netherlands; dDepartment of Pharmaceutical and Pharmacological Sciences, KU Leuven, Leuven, Belgium; eDepartment of Public Health, Ghent University, Ghent, Belgium; fDepartment of Economics, University of La Rioja, Logroño, Spain; gCEReSS - Health Service Research and Quality of Life Center, Aix-Marseille University, Marseille, France; hPublic Health Department - Research Unit, Aix-Marseille University, Marseille, France; iTechnische Hochschule Nürnberg, Nürnberg, Germany; jPharmacotherapy Faculty, College of Pharmacy, University of Utah, Salt Lake City, USA; kThe Swedish Institute for Health Economics (IHE), Lund, Sweden; lNuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK; mGlobal Health Economics and Outcomes Research and Real World EvidenceAveXis Inc, Novartis Gene Therapies, Bannockburn, IL, USA

ABSTRACT

Objective: To provide recommendations for addressing previously identified key challenges in health economic evaluations of Gene Replacement Therapies (GRTs), including: 1) the assessment of clinical effectiveness; 2) the valuation of health outcomes; 3) the time horizon and extrapolation of effects beyond trial duration; 4) the estimation of costs; 5) the selection of appropriate discount rates; 6) the incorporation of broader elements of value; and 7) affordability.

Methods: A literature review on economic evaluations of GRT was performed. Interviews were conducted with 8 European and US health economic experts with experience in evaluations of GRT. Targeted literature reviews were conducted to investigate further potential solutions to specific challenges.

Recommendations: Experts agreed on factors to be considered to ensure the acceptability of historical cohorts by HTA bodies. Existing prospective registries or, if not available, retrospective registries, may be used to analyse different disease trajectories and inform extrapolations. The importance of expert opinion due to limited data was acknowledged. Expert opinion should be obtained using structured elicitation techniques. Broader elements of value, beyond health gains directly related to treatment, can be considered through the application of a factor to inflate the quality-adjusted life years (QALYs) or a higher cost-effectiveness threshold. Additionally, the use of cost-benefit analysis and saved young life equivalents (SAVE) were proposed as alternatives to QALYs for the valuations of outcomes of GRT as they can incorporate broader elements of value and avoid problems of eliciting utilities for paediatric diseases.

Conclusions: While some of the limitations of economic evaluations of GRT are inherent to limited clinical data and lack of experience with these treatments, others may be addressed by methodological research to be conducted by health economists.

Introduction

Gene replacement therapy (GRT) is a novel approach to cure diseases caused by monogenic genetic disorders which offers hope for lifelong improvement to patients suffering from inherited, incurable, and debilitating conditions[1]. Gene-based therapies use genes to treat diseases, but they can employ different techniques to repair a defect in an existing DNA[2]. In GRT, a working gene is delivered to the cell by a vector. This technique conveys a proper gene but does not alter a person's DNA. Once the new gene is present in the nucleus of the cell, it starts to produce the missing protein which is included in metabolic pathways delivering proper metabolites or restoring defective enzymes[3]. The uniqueness of GRT encompasses an innovative and novel approach that offers the potential of a one-time cure.

GRTs have seen dynamic development and, to date, several GRTs have been approved for use in the EU and the US: Zolgensma® – onasemnogene abeparvovec-xioi (US); Zynteglo® – autologous CD34+ cells encoding βA-T87Q-globin gene (EU); Luxturna® – voretigene neparvovec (US/EU); Strimvelis® – autologous CD34+ enriched

cell fraction containing CD34+ cells transduced with a retroviral vector that encodes for the human ADA cDNA sequence (EU); and there are several under study in clinical trials (including CERE-110 – Adeno-Associated Virus Delivery of NGF, AAV-mediated REP1 gene replacement, AAV5-hFIXco-Padua, OTL-103, OTL-200, SGT-53, GS010, VM202) [4,5].

GRTs have the potential to 'cure' diseases. It is important to clarify the meaning of 'cure' in this context. 'Cure' does not necessarily mean that patients will have a normal life, but rather that the underlying pathophysiology will be suppressed. Some irreversible consequences of the disease, occurring before treatment, may persist.

The unique nature of gene therapies has raised major methodological challenges for health-economic assessments [4,6–8]. To date, the following key challenges and considerations have been identified with such analyses: 1) the assessment of clinical effectiveness and safety; 2) the extrapolation of effects beyond trial duration; 3) the valuation of health outcomes; 4) the estimation of costs; 5) the selection of appropriate discount rates; 6) the incorporation of equity considerations; and 7) affordability.

The objective of this paper is to provide potential methodological solutions and recommendations for addressing each of the identified key challenges and considerations, with the aim that future health economic evaluations of GRT will better inform decision-making. The scope of the recommendations focuses on GRTs indicated for serious diseases with high mortality (life expectancy reduced by at least 20% compared with the general population life expectancy) and/or major disability, in situations where no alternative treatment exists, or available treatments have low and unsatisfactory efficacy results. Therapies that replace existing effective treatments are considered beyond our scope, as for haemophilia for example.

Methods

Targeted literature searches were performed in August 2019 using Medline to identify published economic evaluations of GRT and recommendations or discussion papers on the methods of evaluation of gene therapies. In addition, we searched for health technology assessments (HTAs) from the UK National Institute for Health and Care Excellence (NICE) and Institute for Clinical and Economic Review (ICER) for GRT. After identification of challenges associated with the evaluation of GRT, additional target searches were performed in Medline to identify methodological

publications and recommendations related to several issues: assessment of relative effectiveness based on single-arm trials, valuation of health outcomes in children, elements of value beyond QALYs, monetary valuation of caregiver time, elicitation of expert opinion for health economic models, incorporation of equity considerations.

A preliminary list of potential methodological solutions based on the literature reviews was shared with an expert panel comprising 8 US and European health economic experts (from countries including Belgium, France, Germany, Spain, Sweden, the Netherlands and the UK), with experience relevant to the evaluation of GRTs. All experts are co-authors of this manuscript. The experts were asked to review the list of potential solutions, and all comments received were discussed during a board meeting (teleconference). Additional targeted literature reviews were conducted to investigate further issues raised during the meeting, including statistical methods for extrapolation of health outcomes, utilisation of cost-benefit analysis as an alternative to cost-effectiveness analysis and selection of appropriate discount rates. A first version of recommendations was then written and circulated to all co-authors. The co-authors provided comments by email, the comments were summarised, and the recommendations were updated. The new version was circulated together with the summary of comments, and this process was iterated three times until reaching a consensus.

Methodological issues in economic analyses of GRT

There were six publications on economic evaluations of GRT identified in August 2019, including 5 for existing treatments (Luxturna®, Strimvelis® and Zolgensma®) and 1 for a hypothetical haemophilia treatment [9–14]. Five were published in scientific journals, and one was a report from ICER. The key methodological issues in those studies were limitations of clinical data (small clinical trials, unknown duration of treatment effect), lack of appropriate data to extrapolate outcomes in treated patients, limitations of utility data, and sparse data on current practices.

In addition, we identified three recommendation or discussion papers about the economic evaluation of gene therapies [6,7,15]. This included an article by Drummond et al., providing a checklist for analysts and decision-makers to determine which aspects of economic evaluation should be considered further, given the unique nature of gene therapy[15].

Estimating costs

The Second Panel on Cost-Effectiveness in Health and Medicine recommended that all studies represent a reference case based on health care payer perspective and a reference case based on societal perspective[16]. This recommendation appears appropriate in the case of GRTs, as healthcare payers require evaluations from health-care perspective in many countries [17], but costs outside the healthcare sector are particularly high, and would be difficult to ignore. Diseases targeted by GRTs have substantial costs outside of the health-care sector, including social services, special education, and reduced productivity due to disability. Based on an economic burden study conducted in Spain, the direct non-healthcare costs accounted for 67.7% of the total cost of spinal muscular atrophy (SMA) patients (€22,839 out of €33,721 of the total annual cost per SMA patient) [18]. Likewise, Hendrie et al [19]. reported a mean annual cost per case of a Rett Syndrome patient of about 20,000, USD and high-cost items included long-term residential care ($9,371), therapy services out of school, and paid home and community care ($7,213). In addition, caregiver productivity loss and time spent on caregiving are relatively large in the context of diseases targeted by GRT.

It may be noted that these costs will not necessarily decrease following treatment. As noted for Rett syndrome, caring for patients with reduced symptoms could actually take more time than when symptoms are more severe[20]. While social and educational costs would also likely increase as symptoms improve, patients also become more active and, therefore more able to utilize and benefit from available resources. In addition, there will be healthcare costs and other costs, related to the disease or not, during added years of life[21].

As the time of caregivers is to be considered, an important question to address is how to value an unpaid caregiver's time. Two main approaches have been proposed: caregiver time can be valued either based on the income generated by the caregiver if (s) he had been doing paid work instead of caring for a relative, or based on the cost of hiring a professional caregiver for providing the same service[22].

There is no consensus about a preferable approach in the literature. In a recently published systematic review on the valuation of informal care in cost-illness studies, authors highlighted that informal care is not consistently included in the economic studies, and there is a need for greater transparency and clarity in their methods and results[22]. None of the methods (opportunity cost, proxy good and contingent

valuation) might be assessed as the most appropriate. The suggestion for researchers is to use more than one method of valuation to facilitate comparability of studies and increase the usefulness of the analysis for decision-makers [23,24]. Those authors also noted that most OECD countries are moving towards a model of shared responsibility in caring for people with limited autonomy, which means that informal care will be replaced by formal care (for instance, Spain is expected to experience an increase in informal care from approximately 1.2 million persons in 2010 to more than 2.8 million persons in 2060)[23].

In addition, as GRTs may substantially improve life expectancy, the question of whether to include future costs unrelated to the disease arises[25]. Some authorities opt for the inclusion of such costs (e.g. the Second Panel on Cost-Effectiveness in Health and Medicine, The Swedish Dental and Pharmaceutical Benefits Agency [TLV], the National Health Care Institute in the Netherlands [ZIN], Institute for Clinical and Economic Review in the US), while others prefer not to include them (e.g. National Institute for Health and Care Excellence [NICE] in England, Haute Autorité de santé Commission Évaluation Économique et de Santé Publique [HAS CEESP] in France)[26]. There are in fact, strong arguments in favour of including the costs unrelated to the diseases over life-years gained in the evaluation. Essentially, they represent resources expended that are a consequence of the intervention. Van Baal et al. [25] argued that ignoring these costs in a cost-effectiveness analyses is contrary to common sense, results in lost health, and fails to fully inform decision-makers who rely on these analyses. Indeed, there are health benefits associated with those future health expenses, which are implicitly included in evaluations; therefore, considering unrelated future medical costs may be appropriate, provided that the costs are framed in the context of the benefits they represent.

In the case of GRT, this question might be extended to other future costs (e.g. social, educational) which, while significant, are also associated with benefits (e.g. productivity gains) to patients, caregivers, and society at large – and are therefore generally regarded as societal obligation for all children. If a child's life is extended, and that child requires special education during added years of life, it may be appropriate to account for special education costs as they are 'related to the disease'. If the child was able to receive standard education, this may be considered as 'unrelated to the disease' and therefore not considered. Thus, if the chosen approach was to exclude unrelated future 'other' costs, then only the excess costs of special education relative to standard

education should be included, to ensure comparability between products. Based on this reasoning, in the situation where the child's condition is so severe that special education possibilities are limited during lifetime gained, education costs may be lower than standard education, and therefore negative costs of education would be applied. In the comparator arm, where the child dies earlier, the full costs of education should be subtracted. Thus, if education costs are considered relevant to the evaluation, both approaches (including or excluding education costs unrelated to the disease during added years of life) may be acceptable and provide similar results, as long as standard education costs are subtracted for children who do not receive it in the approach where future education costs 'unrelated to the disease' are excluded.

Estimating health outcomes

Assessment of clinical effectiveness

Clinical studies for GRTs frequently present several limitations that may affect economic analysis. Because many GRTs target rare diseases, the sample sizes of clinical studies are small, with some studies having less than 15 patients. A recent study aiming to characterise ATMPs (including GRTs) in development reported that approximately half of the trials (47.2%) enrolled fewer than 25 patients[27].

Another limitation of clinical studies of GRTs is the heterogeneous patient population, because of high interindividual variability in the clinical course of monogenic disorders, typically targeted by GRTs, as well as in the baseline characteristics of patients[28]. Thus, Hanna et al. highlighted in their review that 20% of identified trials recruited both children and adults[27].

Furthermore, surrogate endpoints are frequently used in clinical trials of GRTs. For example, SMA type 1 has many different symptoms, which may appear at different stages of life. The ability to sit was chosen as one of the primary outcomes in phase 3 clinical trial of Zolgensma® because it is an important goal for patients to achieve, and also because it is expected that this milestone correlates with many other symptoms[29]. The use of surrogate outcomes makes it difficult to use clinical research results in economic evaluation, especially in rare and understudied conditions where validation of surrogacy is challenging[15].

The following paragraphs address three key challenges for the assessment of clinical effectiveness for gene therapies: 1) validation of surrogate outcomes; 2) the lack of comparison groups in clinical studies, which may lead to comparisons vs. historical cohorts; and 3) challenges related to the extrapolation of short-term evidence to long-term benefit (discussed below in the «time horizon and extrapolation» section [3.4]).

Validation of surrogate outcomes

Advantages and disadvantages of using surrogate outcomes in clinical trials have been discussed in the literature [30,31]. In cases where employing surrogate outcomes is necessary, information about their validation and evaluation should be considered. Approaches to the validation of surrogate outcomes were discussed by Ciani et al. [30] and includes a 3-level hierarchy approach: (1) biological credibility, (2) determination if any relationships exist between the surrogate and the outcome at the cohort and individual patient level, or (3) evidence from several clinical trials of correlation between treatment effects on surrogate outcomes and final outcomes. Demonstrating the correlation at the individual level for diseases targeted by GRTs may be feasible based on observational studies. However, the correlation between treatments effects surrogate and final health outcomes is not available at the time of the analysis for most GRTs, due to scarcity of data.

Use of historical cohorts

The results from clinical trials conducted on small samples and/or performed without appropriate comparators can, in some instances, be compared with those obtained from previous studies conducted on cohorts similar to the population of interest. Comparison to a historical cohort was used, for example, in the multicentre clinical development program of Luxturna®, an emerging gene therapy for the treatment of biallelic RPE65-mediated inherited retinal disease [32], as well as in its evaluation by the Institute for Clinical and Economic Review[9]. Historical controls may include prior patients with the same disorder from an observational study (prospective natural history study, medical chart data from clinical care), or a control group from a prior randomised investigational study.

To ensure the validity of comparisons to historical observational or investigational cohorts and acceptance by HTA agencies, the following aspects need to be considered [28,33]:

1. The rationale behind not doing a comparative trial is provided, including but not necessarily limited to the argument that this type of studies can increase the risk of irreversible damage, dramatically delaying access to poorly serviced patients and that the recruitment of patients has to be done from very small populations.

2. Preliminary data suggest that the magnitude of the treatment effect size versus the historical cohort is dramatic.

3. The primary endpoint is objective, durable and reproducible.

4. The impact of study heterogeneity in the patient population and the impact on the outcome is studied.

5. Confounding factors affecting the outcome are relatively well known, and a statistically sound adjustment method is used to control for confounding factors.

The use of historical cohorts requires methodological approaches for matching or making indirect comparisons. Collection of equivalent data from historical studies can be difficult, especially when relying on published sources. Berger et al. developed a list of criteria that should be taken into account in the evaluation of the relevance and credibility of observational studies[34]. They are also useful in judging if existing data are suitable for matching and comparisons.

An exact matching method or a propensity score method can be used to ensure that patients are paired on key variables of interest[35].

6. The generalisability and transferability of the clinical data toward the historical cohort should be proactively assessed.

Estimating health-related quality of life

The quality-adjusted life year (QALY) measure has been widely challenged but still remains the most accepted approach for modelling the incremental benefit of health technologies in many countries [36,37]. The incremental cost per life-year gained may in some instances be sufficient to justify the value for money (or lack of) of a GRT, or it may be a useful addition to the incremental cost per QALY gained. However, some consideration of health-related quality of life (HRQoL), with and without GRT, will be required in many evaluations. In the following paragraphs, we consider options to estimate QALYs gained for GRT and alternatives to the QALY measure.

Estimating health-related quality of life in very young children

Many patients targeted by GRTs are likely to be too young, with significant cognitive damages, or otherwise incapable of describing their own HRQoL. Several GRTs in development to date are indicated for diseases affecting very young children (<5 years of age)[38]. While some therapies target diseases affecting older patients, and prevalent cases above 5 years of age may also be treated, it is expected that, in the future, patients may be screened, either systematically or based on the presence of risk factors, and treated before the clinical expression of the disease to prevent the onset of clinical symptoms and irreversible damage. Therefore, proxy reports of patients' HRQoL are likely to be needed[39].

In the case of young children, the most appropriate proxies will likely be parents or caregivers[40]. Proxy reports should be used for observable concepts, but are often unreliable when it comes to concepts that require interpretation, such as social functioning and emotional well-being. Signs and symptoms reported in terms of frequency are likely to be easier to observe and report than those expressed in terms of severity[40]. In addition, proxies' own HRQoL should be collected in order to evaluate the extent to which their own HRQoL impacts the perceived burden on their child[41].

Generic health status classifications such as the EQ-5D [42] or SF-36 [43] are widely used for the evaluation of interventions in adults. However, as noted by several researchers, they may not be appropriate for assessing health states in very young children (<5 years of age) [20,44–46].

Nine preference-based paediatric HRQoL measures have been identified in the literature. They vary across a number of different characteristics, including the number and composition of dimensions, item structures, evaluation protocols, the age groups for which they were validated, and characteristics of target responders [40,47].

Estimating health-related quality of families and caregivers

The burden of caregivers, involving both the burden of caring and the emotional distress associated with the suffering of a close relative, is important to capture in evaluations of GRT. In the study Killian et al [48]. the Optum™ SF-36v2® Health Survey was used to assess Rett Syndrome (RTT) caretakers' physical and mental QOL. RTT characteristics were found to significantly impact HRQOL. Interestingly, more severe disease was associated with a poorer score for the physical component of quality of life (PCS, p = 0.006), but improved the mental component score (MCS, p = 0.003). A study among mothers of children with SMA found a higher mean caregiver burden than studies of caregivers of patients with other neuromuscular diseases. A substantial proportion of mothers (76%) perceived high caregiver burden. Burden, emotional distress and satisfaction with participation were comparable between mothers of children and mothers of adults with SMA[49].

Valuation of health outcomes

Valuing health-related quality of life in very young children

As relevant dimensions of HRQoL change through different stages of child development, the creation of a simple, generic health-status classification system

specific for very young children is challenging. For example, the PedsQL is a very detailed instrument comprising of 23 multiple items and includes four sub-scales of functioning: physical, emotional, social and school. Additionally, the PedsQL Infant Scales composed of 36 items for infants 1–12 months and 45 items for toddlers 13–24 months (each with five sub-scales: physical functioning, physical symptoms, emotional functioning, social functioning and cognitive functioning)[50]. Developing a preference-based index for the PedsQL covering several stages of child development based on a direct elicitation method [e.g. time trade-off (TTO) or standard gamble (SG)] would require shortening it to a maximum of 7–8 items. It would be challenging to perform such a reduction without losing dimensions that are essential at some development stages[51].

In the absence of a relevant generic preference-based instrument that spans the childhood years, we could recommend the use of 'vignette studies', which involve direct elicitation of health states described as vignettes using techniques such as SG or TTO with the general public, parents, or caregivers. The approach of 'vignettes' was used in the context of the evaluation of gene replacement therapies for SMA and for PE65-mediated inherited retinal disease, although in both cases the vignettes were value by clinical experts [52,53]. It is important to minimise investigator bias in the development of vignettes and we would suggest considering guidance on the development of patient-reported outcomes (PRO) instruments for that purpose [54–56].

Furthermore, direct elicitation approaches require highly standardised study protocols to minimise investigator bias. For the TTO, the EuroQol protocols have been considered as a standard[57]. However, these protocols adopt a self-perspective: respondents are asked to imagine that they live in the health states to be valued (e.g. the scripts read as follows: 'you would either live in Life A for Y years and then die, or you would live in Life B for 10 years and then die'). It would probably be very difficult, perhaps even impossible, for adults to imagine themselves in the lives of infants. More generally, the TTO and SG tasks are suitable approaches to elicit utilities from the self-perspective but may not be adapted for valuing health states for others[58]. As a change of perspective for the valuation of health states may be necessary, the person trade-off (PTO) approach would offer a possible solution. Person-trade-off is a technique for valuing health states from a social perspective[59]. It consists of asking people how many outcomes of one kind they consider equivalent in social value to X outcomes of another kind.

Another option could be to value health states in terms of willingness-to-pay, which would imply leaving the cost-effectiveness framework for cost-benefit analysis.

The elicitation of health state utility values is not the only challenge related to the estimation of QALYs for interventions targeting children. The QALY model has been widely challenged [60] and appears to be problematic when it is applied to children. In the QALY model, the value of a health profile, characterised as a sequence of health states, is represented as a sum of products of utility and duration of each health state. It assumes in particular that the value of a health state is independent of its duration and of the health states that come before and after it[21]. However – as an example to illustrate the concept – if we consider a health state characterised by an 'inability to stand', it may be considered normal for an infant below 6 months of age to be unable to stand, and this infant may be assigned a utility of one. If this child remains in a similar state up to the age of two or 3 years, one would probably not consider his/her utility as still equal to one. So, the assumption of mutual independence between quality of life (QoL) and duration of a health state does not hold.

It is possible to consider stratifying the QALY model and eliciting utilities over different periods, representing different stages of 'normal development.' But it remains unclear whether such a stratified model would provide an acceptable approximation of the value of the health profile. Empirical studies would be needed to verify that.

In this context, the saved young life equivalents (SAVE) approach, which has attracted less interest than QALYs in the health economic literature, may be worth considering [21]. The SAVE approach is the main measure that can be used with the PTO approach, mentioned above. The number of SAVEs represents the social value of a treatment outcome relative to a young life being saved. The SAVE approach avoids assumptions of the QALY model such as independence between health state value and duration. Thus, it avoids the need for stratifying valuation tasks according to 'normal' stages of child development. Second, the SAVE values can be elicited using a PTO approach, from a societal perspective, thus avoiding the difficulties mentioned above to elicit utilities from self-perspective.

Valuing health-related quality of life of families and caregivers

In cases where there is evidence of an impact of the disease on the HRQoL of families and caregivers, we recommend taking it into consideration in the QALYs or

another valuation of health outcomes, in line with guidelines from NICE and HAS [61,62]. This should be done irrespective of whether costs are estimated from a healthcare payer or societal perspective.

The use of the QALY model could lead to paradoxical results when applied for caregivers. If we consider the example of a disease with substantial disability and short life expectancy, like SMA type 1, it seems reasonable to imagine that the health state utility of a parent would be moderate during the life of their disabled child, would worsen around the time of death of the child, and then progressively increase to a higher level, as no longer affected by the care burden. Assuming that the impact of mourning on health of parents does not last for a lifetime, the total QALYs for the caregiver may decrease if the child's life is extended, with a similar level of disability. Furthermore, the number of QALYs for the caregiver may decrease even more if the child's condition improves. As mentioned above, MCS of SF-36 was higher in caregivers of patients with more severe forms of Rett syndrome[48]. This raises the question of whether the QALY approach is appropriate to value the impact of GRT for the caregiver. A simple solution would be to exclude QALYs for caregivers when an improvement in child's health is associated with a loss in QALYs for the caregivers. We would also recommend qualitative research to understand how parents feel about seemingly negative effects of their child's health improvement on their life, as this is not sufficiently well understood now, in order to recommend appropriate valuation approaches in the future.

Incorporating broader elements of value in the evaluation of GRT

Lakdawalla et al [63]. list elements of value, beyond health gains, that may need to be considered in the economic evaluation of healthcare interventions. This includes a reduction in uncertainty, fear of contagion, insurance value, value of hope, real-option value, equity, scientific spillovers and disease severity.

The focus here is on gene therapies targeting severe or life-threatening diseases. As defined by Torrance, the QALY approach suggests that 'a gain of equal utility increments anywhere on the scale should be equally preferable for the individual whose utilities are being represented. For example, if an individual's utilities are A: 0.2, B: 0.4, C: 0.6 and D: 0.8, the person should be indifferent to whether the change is from A to B or from C to D'[64]. Taylor et al. aimed to compare increments in the utility of health from different baselines [65] and demonstrated that 'there was a clear tendency to prefer an equal utility increment from an initially more severe

utility of health state compared to a better baseline state.' Similarly, previous studies suggest that society prefers to give priority to patients with a greater proportional shortfall [59,66,67].

The belief here is that another important element of value to consider for GRTs concerns the value of cure. It has been suggested that the value of a cure might be greater than the sum of values of incremental gains representing a similar change in health status overall [68]. This could have importance for some GRTs, which are potentially curative for certain diseases.

Further research is needed to assess the value of cure. While recent evidence from a discrete choice experiment suggests that people do not attach any value to cure per se[69], results to such research questions regarding valuation of a cure may be sensitive to the way the problem is framed.

In the context of gene therapy, Drummond et al. [15,37] also consider scientific spillovers as a relevant broader element of value. Scientific spillovers relate to the impact of a new technology on future generations of patients. The knowledge acquired during the development of a treatment might lead to other more valuable drugs in the future, even to treat very different diseases. However, there are already many companies working on the development of GRTs, and very soon the GRTs under evaluation will no longer generate substantial scientific spillovers. Therefore, we would not consider the research on the value of scientific spillovers as a priority.

Broader elements of value could be taken into consideration in the cost/QALY evaluation framework through some modifiers, such as the application of a factor to inflate the QALYs or a higher cost-effectiveness threshold. Several agencies use higher thresholds under specific conditions, which would likely apply in the case of GRT. Thus, in interim guidelines for the evaluation of highly specialised technologies, NICE indicates that treatments costing up to £100,000 might be considered cost-effective. Above that threshold, NICE would consider whether the incremental cost-effectiveness ratio (ICER) falls within the £100,000 limit after applying weights from 1 to 3 to large QALY gains (≥10)[70]. In Sweden, both TLV and The New Therapies (NT) Council are also accepting a higher ICER for severe conditions. If a condition is severe and ultra-rare, they could accept an ICER of up to €200,000 per QALY[71].

Drummond et al. [15,37] highlight that in practice, HTA agencies consider those elements of value through a deliberative process, and it would be important to identify all relevant elements when presenting evaluations of GRT. Multiple Criteria Decision Analysis (MCDA)

would be another possible approach to include broader elements of value associated with GRT, in a more transparent manner than through deliberation. However, the use of an MCDA approach has been rejected by some HTAs due to key limitations such as being 'entirely mechanistic,' or ignoring opportunity costs[72]. Again, the SAVE approach and cost-benefit analyses would have advantages as elements of value such as cure, hope and insurance would be endogenous to the evaluation.

Additional comments about alternatives to the QALY: SAVEs and cost-benefit analysis

We have suggested SAVEs and cost-benefit analysis could be potential alternatives to overcome some challenges related to the estimation of QALYs in the evaluation of GRT. It is important to also address the limitations of these two approaches.

One may question the practical aspects of generating valuations for every relevant health profile using the SAVE approach. If SAVEs were used in the context of a state-transition model, values would be needed for a large number of possible combinations of health states and health-state durations. However, several health profiles could be presented in a structured way to respondents, and a regression analysis could then be used to assess the impact of different dimensions or characteristics of a health profile on its value and generate an algorithm to predict health profile values. It should be acknowledged that SAVEs cannot replace QALYs in all situations. SAVEs would likely be suitable for the evaluation of curative treatments (such as GRT), whereas QALYs would be more suitable for treatments with smaller health gains. In addition, we recommend the conduct of further empirical studies to assess the validity of a SAVE approach as a way to value lifetime health profiles and to support the development of standard protocols to apply this approach.

Cost-benefit analysis could also solve some of the problems associated with QALYs (valuation of health for very young children and incorporation of broader elements of value), but would be a further departure from the cost-utility analysis. The notion of valuing life and quality of life in monetary terms is challenging for clinicians and many healthcare decision-makers. The utilisation of SAVEs would be compatible with an extra-welfarist framework: units of health outcomes would be valued equally for all. This is not the case with CBA, in which the valuation of outcomes would be influenced by the ability to pay[73]. Furthermore, there is a lack of consensus on methods to elicit willingness-to-pay. Results may vary substantially depending on the method used and the

framing of the question[74]. Finally, this is still to be studied empirically, but it might be easier for members of the general public to value large health gains in terms of SAVEs rather than in terms of money.

Selection of an appropriate time horizon and extrapolation methods

Selection of an appropriate time horizon

GRTs are expected to have lifetime consequences and/or extend life expectancy. Therefore, costs and outcomes should be projected over lifetime according to general recommendations for choice of the time horizon [75,76]. However, some recent HTA guidelines stress that the choice of the time horizon is primarily a matter of arbitration between the information generated in retaining a time horizon sufficiently long to integrate all incremental cost and outcomes, and the uncertainty that a time extrapolation generates[77]. While a lifetime horizon may seem desirable for GRTs, it may be misleading for decision-makers if there is no way to know whether the net benefits of treatment will be positive or negative in distant years.

One solution to palliate long-term uncertainty would be to assess scenario analyses with different time horizons pertaining to different insights on the treatment benefit. However, when different scenarios produce a wide range of ICERs, some expert guidance would be needed for decision-makers to weigh the different results presented to them. Without such guidance, risk-averse decision-makers might be biased towards analyses over shorter time-horizons, with less uncertainty, and others might be biased towards analysis over longer time-horizons, potentially capturing all incremental costs and benefits, although highly uncertain. It is therefore suggested to involve a range of experts, through a Delphi panel, to estimate the likelihood of persistence of treatment benefits over different time-horizons, rather than arbitrarily choosing a time horizon. The panels should include disease specialists and geneticists.

Extrapolation methods

The limited duration of clinical trials may not be sufficiently long to measure the impact of treatment on all symptoms, and a challenge will be to determine which consequences of the disease are reversible and which are not. Furthermore, it may take time for patients to develop abilities that were not accessible to them before treatment, and the development of such abilities will also need to be extrapolated.

Standard modelling techniques used in economic evaluations, such as Markov models and Discrete Event Simulation, will likely be appropriate for GRT. The challenge will be to find appropriate data, such as transition probabilities, to populate these models. When there is uncertainty around the proportion of cured patients, then mixture cure models may be helpful to determine the probability of reaching key development endpoints[78]. Let us take the example of a neuromuscular disorder, where relevant health states would be the 'ability to sit' and the 'ability to stand', and the endpoint of the trial is the 'ability to sit'. A first step would be to extrapolate the probability of developing the ability to sit. The observation of the Kaplan–Meier curve for time to the development of the ability to sit will indicate whether this curve appears to be converging towards a plateau or not (i.e. whether or not some patients will remain unable to sit for all their life). If it appears that the curve converges towards a plateau (i.e. that a proportion of patients will never reach the ability to sit, then a mixture-cure model may be used to extrapolate the probability of the patient developing the ability to sit).

Continuing with the same example, another data source will be required for the estimation of the probability of transition to standing. The transition probability observed in children without a disease will not be a relevant estimate in many cases. Probabilities estimated from historical data, perhaps from children with a less severe form of the disease, who reached the ability to stand without treatment, maybe a more relevant proxy. Thus, a possible extrapolation approach would be to assess if the progression of the disease in treated patients matches the progression observed in historical patients with a less severe form of the disease, and, if so, use the information from those historical patients to extrapolate.

Eliciting expert opinion

Clearly, even one makes the best of available data, clinical expert opinion will be required to inform such decisions as to which population to use as a proxy to estimate the long-term transition probabilities for cured patients or which consequences of the disease are reversible. It will also be required to assess the risk of long-term serious adverse events of GRT.

While methodological guidance on elicitation procedures for HTA is needed [56], the elicitation literature recommends a number of good practice steps to be taken when conducting expert elicitation to minimize bias[79]. This includes providing to experts training for the session, formal background of the session, getting feedback from the experts and proposing the possibility of amending their inputs, up to date. Several elicitation procedures are available to obtain information from experts and make a probabilistic representation of their knowledge. Two different approaches of elicitation are commonly used in the literature of structured elicitations for cost-effectiveness analyses: 1) the fixed interval method, in which the expert reports his/her probability of the uncertain quantity of interest θ, for example the recurrence rate, the mortality rate or the time to death, lying in specified intervals, and 2) the variable interval method, in which he/she makes quantile judgements[80]. Among the common fixed interval elicitation methods, is the trial roulette method, also called the 'chips and bins method', where the expert provides probabilities of θ lying in a particular 'bin' by allocating 'chips' to that bin [81,82]. More recently, a five-step method has been developed, suggesting decomposing the elicitation task in five steps, including more feedback in particular, to improve the quality of the probability distributions that results from the elicitation[83].

Selection of appropriate discount rates

The choice of an appropriate uniform discount rate for GRTs is part of the public sector investment strategy. The rate at which future costs and benefits of publicly funded programmes should be discounted is called a 'social discount rate' and is defined by either 'time preference' or 'opportunity costs' rates. Opportunity costs measure the value to the society of the next best alternative use to which funds employed in a public project might otherwise have been put, taking into account the possible market distortions. The opportunity costs of capital are estimated by the pre-tax marginal rate of return on private investments observed in the marketplace[84]. Time preference measures the rate at which society is willing to trade current benefits and costs for its future values. The time preference rate is based on the tax-free rate of return on government bonds or other low-risk market securities[19].

The appropriate use of discounting is critical for models that represent a long-time horizon due to the compounding effect of discounting. For GRT, the treatment cost occurs upfront, but the benefits accrue over a lifetime. Thus, the choice of the discount rate has a large effect on estimates of cost-effectiveness[85]. The use of a lower discount rate for health outcomes would generally lead to lower ICERs for GRTs compared to no treatment[86].

The recommended discount rate varies across HTA agencies, ranging from 1.5% for some Nordic countries to 5% for commercial payers in the US. It is common to

apply a uniform discount rate to both costs and outcomes and keep it constant over time, in line with recommendations from most agencies including NICE [61] and the Washington Panel on Cost Effectiveness in Health and Medicine[87]. Interim process and methods of the highly specialised technologies programme (2017), published by NICE, specify that a non-reference-case discount rate of 1.5% (instead of 3.5%) for costs and benefits may be considered when 'treatment restores to full or near full health, who would otherwise die or have impaired lives, when benefits will be sustained for normally at least 30 years … '. Unlike in most countries, the recommended discount rate varies between costs and benefits in Belgium, with 3% for costs and 1.5% for benefits [88], and in the Netherlands, with 4% for costs and 1.5% for benefits[89].

There is controversy about whether costs and benefits should be discounted at the same rate. Uniform discounting is supported by two main arguments: the consistency thesis and the postponement paradox. The consistency thesis proposes that inconsistencies may occur when discounting at two different rates[90]. The postponement paradox, presented by Keeler and Cretin, is that if health benefits are discounted at a lower rate than costs, the cost-effectiveness ratio can be improved by delaying the introduction of the technology in question and continue to be improved by further delays[91]. However, this argument may not be relevant as decision-makers typically choose between competing priorities to fund from a fixed budget, rather than the optimal timing of the health technology. Most recently Claxton et al [92]. argued that the soundness of differential discounting depends on whether the decision-maker is seeking to maximise welfare or health; whether the budget for healthcare is fixed; whether the value of health changes over time; and determined social time preference rates. Authors suggest that differential discounting should take place only if the marginal productivity of health spending changes over time[84].

There are strong reasons to argue that discount rates generally recommended by HTA agencies are too high, in particular for health outcomes, and this is an important consideration for the evaluation of GRTs. First, Arrow and Lind argued that the relevant required rate of return on public investment is the risk-free interest rate[93]. This can be approximated by the interest rates on government bonds of 10 years or less, which are currently below 1% in the US and EU countries[94]. In addition, empirical evidence suggests that the real interest rates around the world have come down for both private and public markets, and the new theoretical advances considering future uncertainty likely suggest lower long-term rates as well. The council of Economic Advisers (2017) provides evidence supporting lowering these discount rates to, at most, 2%[84]. Furthermore, the monetary value of health (i.e. the rate at which we exchange consumption for health, our 'willingness to pay' for health) is expected to grow with increases in income over time [95–97]. This can be accounted for by reducing the discount rate for health outcomes.

Finally, it is also relevant to consider recommendations for the evaluations of vaccines, which present the same characteristics as GRTs, i.e. high upfront costs and health benefits spreading over the long term. A consensus framework from a European vaccines economic community advised that the discount rate for health effects should be around half of the discount rate for costs[98]. The WHO even advocates a 0% discount rate for health effects of vaccination[99].

Affordability and new payment models

Cost-effectiveness analysis is a method to maximise QALYs (or health gains) under a budget constraint. Assuming there is a fixed budget constraint, the theoretical solution would be to lower the cost-effectiveness threshold when the budget is exceeded. However, budget constraints may be set for periods of 1 to 5 years, and cost-effectiveness analysis does not account for the distribution of costs (and benefits) over time. With GRT, there would be potentially high upfront drug acquisition costs, and cost offsets in the long term. Thus, assuming everything else equal, the budget constraint may be exceeded in the coming years where many GRT therapies may be launched simultaneously, and costs may be below the budget constraint in distant years. Does it mean that GRT should be considered as an investment for the future, and payers should borrow money to pay for GRT? So far in public accounting this is not possible, and the law is required to allow for depreciation through intangible amortisation of specific pharmaceuticals such as GRTs. Should we look for ways to account for the distribution of costs and benefits in a cost-effectiveness analysis? While researchers investigate this question, budget impact analysis will be particularly important to inform healthcare payers making decisions regarding GRT.

Summary

Main challenges	Recommendations	Sources
Assessment of clinical effectiveness based on small clinical trials, often single-arm	• The results from clinical trials conducted on small samples and/or performed without appropriate comparators can in some instances be compared with those obtained from previous studies conducted on cohorts similar to the population of interest. Historical controls may include prior patients with the same disorder from an observational study (prospective natural history study, medical chart data from clinical care), or from a control group from a prior randomised investigational study.	• Literature review[a]
	• Factors to be considered to ensure historical cohorts acceptability by HTA bodies: (1) The rationale behind not doing a comparative trial is provided. (2) Preliminary data suggest that the magnitude of treatment effect size versus historical cohort is dramatic. (3) The primary endpoint is objective, durable and reproducible. (4) The impact of study heterogeneity in the patient population and the impact on the outcome is studied. (5) Confounding factors affecting the outcome are relatively well known, and a statistically sound adjustment method is used to control for confounding. (6) The generalisability and transferability of the clinical data toward the historical cohort are proactively assessed.	• Augustine 2013[28], Rémuzat 2019[33]
Estimation of costs	• Costs should be evaluated from a health care perspective and a societal perspective	• Sanders 2016[16]
	• Caregiver time can be valued either based on the income generated by the caregiver if (s)he had been doing paid work instead of caring for a relative, or based on the cost of hiring a professional caregiver for providing the same service.	• Oliva-Moreno 2017 [23], Hjortsberg 2010 [24]
	• Including future unrelated medical costs, along with associated health benefits (which may be considered implicitly), would be appropriate.	• Van Baal 2019[25]

Valuation of health outcomes	→ **Valuing quality of life in very young children**
	• The Saved Young Life Equivalents (SAVE) approach, which has attracted less interest than QALYs in the health economic literature, may be worth reconsidering.
	  ○ First, the SAVE approach would avoid assumptions of the QALY model such as independence between health state value and duration, which is not sustainable.
	  ○ Second, SAVEs would be elicited using a Person Trade-off (PTO) approach, from a societal perspective, thus avoiding the difficulties to elicit utilities for very young children from self-perspective.
	• A cost-benefit analysis would avoid the problems related to the elicitation of utilities for very young children, but the valuation of quality of life in monetary terms is challenging for clinicians and many healthcare decision-makers.
		Nord 1992[21], panel[aa]

	→ **Valuing the HRQoL of families and caregivers** • The burden of caregivers, involved by both emotional distresses facing suffering from a disease of a close relative as well as by the burden of caring, will be substantial considering the severity of diseases treated by GRT. • When there is evidence of an impact of the disease on the HRQoL of families and caregivers, this should be accounted for in the evaluation of GRTs, irrespective of whether the analysis is performed from a healthcare payer's or societal perspective.	Review[a], NICE[61], HAS[62]

(Continued)

(Continued).

Main challenges	Recommendations	Sources
	→ **Incorporating broader elements of value** • Broader elements of value could be taken into consideration in the cost/QALY evaluation framework through some modifiers, such as the application of a factor to inflate the QALYs or a higher cost-effectiveness threshold. • Drummond et al[15,37]. Highlighted the fact that in practice, HTA agencies consider those elements of value through a deliberative process, and it would be important to identify all relevant elements when presenting evaluations of GRT.	• NICE 2019[61], Medic 2017[71] • Drummond 2013[37], Drummond 2019[15]
Time horizon and extrapolation, as there is substantial uncertainty around long-term effects, positive or negative	→ **Selection of an appropriate time horizon** • While a lifetime horizon may seem desirable for GRT, it may be misleading for decision-makers if we have no way to know whether the net benefits of treatment will be positive or negative in distant years. • One solution to palliate long-term uncertainty would be to assess scenario analyses with different time horizons pertaining to different knowledge about treatment benefit. • However, when different scenarios produce a wide range of ICERs, some expert guidance via the use of Delphi panels could be useful for decision-makers to weigh the different results presented to them.	**Panel****
	→ **Eliciting expert opinion** • Experts acknowledged the importance of expert opinion due to limited data available in the context of GRT. • Several elicitation procedures are available to obtain information from experts and make a probabilistic representation of their knowledge. • Two different approaches of structured elicitation are recommended for cost-effectiveness analyses: 1) the fixed interval method, in which the expert reports his/her probability of the uncertain quantity of interest θ, for example the recurrence rate, the mortality rate or the time to death, lying in specified intervals, and 2) the variable interval method, in which (s)he makes quantile judgements.	Panel**, Grigore 2016[79], Soares 2018[80]
	→ **Extrapolation methods** • Standard modelling techniques used in economic evaluation, such as Markov models and Discrete Event Simulation, will likely be appropriate for GRT. The challenge will be to find appropriate data, such as transition probabilities, to populate these models. • Information from historical patients might be used to generate the input data. • When there is uncertainty around the proportion of cured patients, then mixture cure models may be helpful to determine the probability of reaching key development endpoints.	• Panel**, Hettle 2017 [7] • Literature review[a] • Yu 2008[78]
Discount rate	• There are strong reasons to argue that discount rates recommended by many HTA agencies are too high, in particular for health outcomes. This is an important consideration for the evaluation of GRTs, as higher discount rates may lead to substantially higher ICERs.	• Klock 2005[95], Brouwer 2005[96]

[a]Recommendations based on the review of published evaluations of GRT; **Recommendations based on expert panel meeting

Conclusion

Some of the limitations of economic evaluations of GRT are inherent to limited clinical data and lack of experience with GRT. Even if uncertainty around evaluation of results cannot be avoided, the use of appropriate methods such as matched comparisons to historical cohorts, extrapolation using mixture models, and structured expert elicitation may help to make analyses more useful for decision-makers. In addition, methodological research would be useful to further assess the potential of methods such as SAVEs and cost-benefit analysis to improve the quality of evaluations of GRTs, as well as to determine appropriate discount rates.

Disclosure statement

FUF reports personal fees and non-financial support from Novartis Gene Therapies (formerly AveXis, Inc.), outside the submitted work. DM reports personal fees from Novartis Gene Therapies (formerly AveXis, Inc.), outside the submitted work. OD is an employee of Novartis Gene Therapies (formerly AveXis, Inc.), a company which commercialises gene therapies. MP reports grants and personal fees from various pharmaceutical industries, all outside the submitted work. He holds stocks in Ingress Health and Pharmacoeconomics Advice Groningen (PAG Ltd) and is an advisor to Asc Academics, all pharmacoeconomic consultancy companies. SA, KT, CF, AM, and MT are employees of Creativ-Ceutical or were employed by Creativ-Ceutical at the time of the development of the guidelines. Creativ-Ceutical is a consulting company in the field of health economics. All other authors report no conflict of interest.

ORCID

Maarten Postma http://orcid.org/0000-0002-6306-3653
Mondher Toumi http://orcid.org/0000-0001-7939-7204

References

[1] Sun W, Zheng W, Simeonov A. Drug discovery and development for rare genetic disorders. Am J Med Genet Part A. 2017;173(9):2307–2322. [published Online First: 2017/07/22].

[2] Boulad F, Mansilla-Soto J, Cabriolu A, et al. Gene therapy and genome editing. Hematol Oncol Clin North Am. 2018;32(2):329–342. published Online First: 2018/02/21.

[3] Keller AS, Iv TCS K, DeLalio LJ, et al. Replacement therapies in metabolic disease. Curr Pharm Biotechnol. 2018;19(5):382–399. [published Online First: 2018/06/21].

[4] Hampson G, Towse A, Pearson SD, et al. Gene therapy: evidence, value and affordability in the US health care system. J Comp Eff Res. 2018;7(1):15–28. [published Online First: 2017/11/17].

[5] Dunbar CE, High KA, Joung JK, et al. Gene therapy comes of age. Science (New York, NY). 2018;359(6372): eaan4672. [published Online First: 2018/01/13].

[6] Jonsson B, Hampson G, Michaels J, et al. Advanced therapy medicinal products and health technology assessment principles and practices for value-based and sustainable healthcare. Eur J Health Econ. 2019;20(3):427–438. [published Online First: 2018/09/20].

[7] Hettle R, Corbett M, Hinde S, et al. The assessment and appraisal of regenerative medicines and cell therapy products: an exploration of methods for review, economic evaluation and appraisal. Health Technol Assess. 2017;21(7):1–204. [published Online First: 2017/03/01].

[8] Marsden G, Towse A Exploring the assessment and appraisal of regenerative medicines and cell therapy products: is the NICE approach fit for purpose? 2017. [cited May 2019]. Available from: https://www.ohe.org/publications/exploring-assessment-and-appraisal-regenerative-medicines-and-cell-therapy-products#

[9] Zimmermann M, Lubinga SJ, Banken R, et al. Cost utility of voretigene neparvovec for biallelic RPE65-mediated inherited retinal disease. Value Health. 2019;22(2):161–167.

[10] South E, Cox E, Meader N, et al. Strimvelis® for treating severe combined immunodeficiency caused by adenosine deaminase deficiency: an evidence review group perspective of a NICE highly specialised technology evaluation. Pharmacoecon Open. 2019;3(2):151–161.

[11] Johnson S, Buessing M, O'Connell T, et al. Cost-effectiveness of voretigene neparvovec-rzyl vs standard care for RPE65-mediated inherited retinal disease. JAMA Ophthalmol. 2019;137(10):1115–1123. [published Online First: 2019/07/19].

[12] Malone DC, Dean R, Arjunji R, et al. Cost-effectiveness analysis of using onasemnogene abeparvocec (AVXS-101) in spinal muscular atrophy type 1 patients. J Mark Access Health Policy. 2019;7(1):1601484.

[13] Machin N, Ragni MV, Smith KJ. Gene therapy in hemophilia A: a cost-effectiveness analysis. Blood Adv. 2018;2 (14):1792–1798.

[14] ICER. Voretigene neparvovec for biallelic RPE65-mediated retinal disease: effectiveness and value; 2017. [cited 2019 Jun 6]. Available from: https://icer-review.org/wp-content/uploads/2017/06/MWCEPAC_VORETIGENE_DRAFT_EVIDENCE_REPORT_11152017.pdf

[15] Drummond MF, Neumann PJ, Sullivan SD, et al. Analytic considerations in applying a general economic evaluation reference case to gene therapy. Value Health. 2019;22(6):661–668. [published Online First: 2019/06/15].

[16] Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. JAMA. 2016;316 (10):1093–1103.

[17] ICER. ICER's reference case for economic evaluations: principles and rationale; 2018. [cited 2019 Jun 6].

Available from: https://icer-review.org/wp-content/uploads/2018/07/ICER_Reference_Case_July-2018.pdf

[18] Lopez-Bastida J, Pena-Longobardo LM, Aranda-Reneo I, et al. Social/economic costs and health-related quality of life in patients with spinal muscular atrophy (SMA) in Spain. Orphanet J Rare Dis. 2017;12(1):141. [published Online First: 2017/08/20].

[19] Hendrie D, Bebbington A, Bower C, et al. Measuring use and cost of health sector and related care in a population of girls and young women with Rett syndrome. Res Autism Spectrum Disorders. 2011;5(2):901–909.

[20] Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. Pediatrics. 2005;115(5):e600–14. [published Online First: 2005/05/04].

[21] Nord E. An alternative to QALYs: the saved young life equivalent (SAVE). BMJ. 1992;305(6858):875–877.

[22] Brouwer W, Rutten F, Koopmanschap M. Costing in economic evaluations. In: Drummond M, McGuire A, editors. Economic evaluation in health care: merging theory with practice. Oxford: Oxford University Press; 2002. p. 68–93.

[23] Oliva-Moreno J, Trapero-Bertran M, Pena-Longobardo LM, et al. The valuation of informal care in cost-of-illness studies: a systematic review. PharmacoEconomics. 2017;35 (3):331–345. [published Online First: 2016/11/17].

[24] Hjortsberg C, Persson U. The value of informal caregiver time for psychotic illness. J Ment Health Policy Econ. 2010;13(3):127–133. [published Online First: 2010/11/06].

[25] van Baal P, Morton A, Meltzer D, et al. Future unrelated medical costs need to be considered in cost effectiveness analysis. Eur J Health Econ. 2019;20(1):1–5. [published Online First: 2018/04/20].

[26] De Vries LM, Van Baal PHM, Brouwer WBF. Future costs in cost-effectiveness analyses: past, present, future. PharmacoEconomics. 2019;37(2):119–130. [published Online First: 2018/11/27].

[27] Hanna E, Remuzat C, Auquier P, et al. Advanced therapy medicinal products: current and future perspectives. J Mark Access Health Policy. 2016;4(1):31036. [published Online First: 2016/04/29].

[28] Augustine EF, Adams HR, Mink JW. Clinical trials in rare disease: challenges and opportunities. J Child Neurol. 2013;28(9):1142–1150. [published Online First: 2013/09/10].

[29] Gene replacement therapy clinical trial for patients with spinal muscular atrophy Type 1 (STR1VE) [updated Nov 2019; cited Nov 2019]. Available from: https://clinicaltrials.gov/ct2/show/NCT03306277

[30] Ciani O, Buyse M, Drummond M, et al. Time to review the role of surrogate end points in health policy: state of the art and the way forward. Value Health. 2017;20 (3):487–495. [published Online First: 2017/03/16].

[31] Vinden C. Surrogate end points save lives. Can J Surg J Canadien De Chirurgie. 2017;60(2):81–82. [published Online First: 2017/03/25].

[32] Zimmermann MR, Lubinga SJ, Rind D, et al. A cost-effectiveness analysis of voretigene neparvovec for vision loss due to biallelic Rpe65-mediated inherited retinal disease. Value Health. 2018;21:S205–S06.

[33] Rémuzat C, Thokagevistk K, Millier A, et al. Factors to be considered to ensure acceptability of historically controlled studies by HTA bodies. ISPOR EU. Copenhagen; 2019.

[34] Berger ML, Martin BC, Husereau D, et al. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC good practice task force report. Value Health. 2014;17(2):143–156. [published Online First: 2014/03/19].

[35] Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1–21.

[36] Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull. 2010;96 (1):5–21. [published Online First: 2010/11/03].

[37] Drummond M, Tarricone R, Torbica A. Assessing the added value of health technologies: reconciling different perspectives. Value Health. 2013;16(1Suppl):S7–13. [published Online First: 2013/01/18].

[38] FDA approves innovative gene therapy to treat pediatric patients with spinal muscular atrophy, a rare disease and leading genetic cause of infant mortality; 2019. [cited November 2019]. Available from: https://www.fda.gov/news-events/press-announcements/fda-approves-innovative-gene-therapy-treat-pediatric-patients-spinal-muscular-atrophy-rare-disease

[39] Hays RM, Valentine J, Haynes G, et al. The seattle pediatric palliative care project: effects on family satisfaction and health-related quality of life. J Palliat Med. 2006;9 (3):716–728. [published Online First: 2006/06/07].

[40] Germain N, Aballéa S, Toumi M. Measuring the health-related quality of life in young children: how far have we come? J Mark Access Health Policy. 2019;7 (1):1618661.

[41] U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes. 2006;4:79. Published 2006 Oct 11. doi:10.1186/1477-7525-4-79.

[42] The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective - 2003 Evidence from the EuroQol BIOMED Research Programme. Springer, 2003.

[43] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002;21(2):271–292. [published Online First: 2002/04/10].

[44] Ungar WJ. Challenges in health state valuation in paediatric economic evaluation: are QALYs contraindicated? PharmacoEconomics. 2011;29(8):641–652. [published Online First: 2011/05/25].

[45] Kind P, Klose K, Gusi N, et al. Can adult weights be used to value child health states? Testing the influence of perspective in valuing EQ-5D-Y. Qual Life Res. 2015;24 (10):2519–2539. [published Online First: 2015/04/22].

[46] Noyes J, Edwards RT. EQ-5D for the assessment of health-related quality of life and resource allocation in children: a systematic methodological review. Value Health. 2011;14(8):1117–1129. [published Online First: 2011/12/14].

[47] Chen G, Ratcliffe JA. Review of the development and application of generic multi-attribute utility instruments

for paediatric populations. PharmacoEconomics. 2015;33 (10):1013–1028. [published Online First: 2015/05/20].

[48] Killian JT Jr., Lane JB, Lee HS, et al. Caretaker quality of life in rett syndrome: disorder features and psychological predictors. Pediatr Neurol. 2016;58:67–74. [published Online First: 2016/03/21].

[49] Cremers CH, Fischer MJ, Kruitwagen-van Reenen ET, et al. Participation and mental well-being of mothers of home-living patients with spinal muscular atrophy. Neuromuscul Disord. 2019;29(4):321–329. [published Online First: 2019/04/08].

[50] Kruse S, Schneeberg A, Brussoni M. Construct validity and impact of mode of administration of the PedsQL™ among a pediatric injury population. Health Qual Life Outcomes. 2014;12(1):168.

[51] Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the pediatric quality of life inventory. Med Care. 1999;37 (2):126–139. [published Online First: 1999/02/19].

[52] Lloyd AJ, Thompson R, Gallop K, et al. Estimation of the quality of life benefits associated with treatment for spinal muscular atrophy. Clinicoecon Outcomes Res. 2019;11:615–622.

[53] Lloyd A, Piglowska N, Ciulla T, et al. Estimation of impact of RPE65-mediated inherited retinal disease on quality of life and the potential benefits of gene therapy. Br J Ophthalmol. 2019;103(11):1610–1614. [published Online First: 2019/01/20].

[54] Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity–establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1–eliciting concepts for a new PRO instrument. Value Health. 2011;14(8):967–977. [published Online First: 2011/12/14].

[55] US Department of Health and Human Services. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims; 2009. [cited November 2019]. Available from: https://www.fda.gov/media/77832/download

[56] European Medicines Agency. Committee for medicinal products for human use. Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products. [cited November 2019.]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf

[57] Oppe M, Rand-Hendriksen K, Shah K, et al. EuroQol protocols for time trade-off valuation of health outcomes. PharmacoEconomics. 2016;34(10):993–1004. [published Online First: 2016/04/17].

[58] Dolan P. Chapter 32 The measurement of health-related quality of life for use in resource allocation decisions in health care. Handbook of Health Economics. Amsterdam: Elsevier Science Bv; 2000. p. 1723–1760.

[59] Nord E. The person-trade-off approach to valuing health care programs. Med Decis Making. 1995;15(3):201–208. [published Online First: 1995/07/01].

[60] Nord E, Daniels N, Kamlet M. QALYs: some challenges. Value Health. 2009;12(Suppl 1):S10–5. [published Online First: 2009/03/11].

[61] NICE. Guide to the methods of technology appraisal. Process and methods [PMG9]; 2013.

[62] HAS. Choices in methods for economic evaluation. A methodological guide; 2012. [cited 2019 Jun 6]. Available from: https://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf

[63] Lakdawalla DN, Doshi JA, Garrison LP Jr., et al. Defining elements of value in health care-a health economics approach: an ISPOR special task force report [3]. Value Health. 2018;21(2):131–139. [published Online First: 2018/02/27].

[64] Torrance GW. Utility measurement in healthcare: the things I never got to. PharmacoEconomics. 2006;24 (11):1069–1078. [published Online First: 2006/10/28].

[65] Taylor M, Chilton S, Ronaldson S, et al. Comparing increments in utility of health: an individual-based approach. Value Health. 2017;20(2):224–229. [published Online First: 2017/02/27].

[66] Ubel PA. How stable are people's preferences for giving priority to severely ill patients? Soc Sci Med. 1999;49 (7):895–903. [published Online First: 1999/09/01].

[67] Reckers-Droog VT, van Exel NJA, Brouwer WBF. Looking back and moving forward: on the application of proportional shortfall in healthcare priority setting in the Netherlands. Health Policy. 2018;122(6):621–629. [published Online First: 2018/04/29].

[68] Husereau D. How do we value a cure? Expert Rev Pharmacoecon Outcomes Res. 2015;15(4):551–555. [published Online First: 2015/04/29].

[69] Hampson G, Mott D, Devlin N, et al. Public preferences for health gains and cures: a discrete choice experiment. OHE Consulting Ltd; 2019. [cited 2019 Jun 6]. Available from: https://www.ohe.org/publications/public-preferences-health-gains-and-cures-discrete-choice-experiment#

[70] NICE and NHS England consultation on changes to the arrangements for evaluating and funding drugs and other health technologies assessed through NICE's technology appraisal and highly specialised technologies programmes; 2017. [cited December 2019]. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/board-paper-TA-HST-consultation-mar-17-HST-only.pdf

[71] Medic G, Korchagina D, Young KE, et al. Do payers value rarity? An analysis of the relationship between disease rarity and orphan drug prices in Europe. J Mark Access Health Policy. 2017;5(1):1299665. [published Online First: 2017/05/06].

[72] Baltussen R, Marsh K, Thokala P, et al. Multicriteria decision analysis to support health technology assessment agencies: benefits, limitations, and the way forward. Value Health. 2019;22(11):1283–1288. [published Online First: 2019/11/12].

[73] Birch S, Donaldson C. Valuing the benefits and costs of health care programmes: where's the 'extra' in extra-welfarism? Soc Sci Med. 2003;56(5):1121–1133. [published Online First: 2003/02/21].

[74] EurVaQ. European value of a quality adjusted life year. Instrument: specific targeted research project. Final publishable report. Sixth framework programme. [cited November 2019]. Available from: https://research.ncl.ac.uk/eurovaq/EuroVaQ_Final_Publishable_Report_and_Appendices.pdf

[75] Drummond MF, Sculpher MJ, Claxton K, et al. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2015.

[76] Peter J Neumann; Theodore G Ganiats; Louise B Russell; Gillian D Sanders; Joanna E Siegel. Cost-effectiveness in health and medicine. 2nd ed. New York: Oxford University Press; 2017.

[77] Choix méthodologiques pour l'évaluation de l'efficience à la HAS: version soumise à la consultation publique – HAS (11 juin 2019); 2019. [cited November 2019]. Available from: http://www.ors-auvergne.org/veille-sante-social/choix-methodologiques-levaluation-de-lefficience-a-has-version-soumise-a-consultation-publique-has-11-juin-2019/

[78] Yu B, Peng Y. Mixture cure models for multivariate survival data. Comput Stat Data Anal. 2008;52(3):1524–1532.

[79] Grigore B, Peters J, Hyde C, et al. A comparison of two methods for expert elicitation in health technology assessments. BMC Med Res Methodol. 2016;16(1):85.

[80] Soares MO, Sharples L, Morton A, et al. Experiences of structured elicitation for model-based cost-effectiveness analyses. Value Health. 2018;21(6):715–723. [published Online First: 2018/06/19].

[81] Soares MO, Bojke L, Dumville J, et al. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. Stat Med. 2011;30(19):2363–2380.

[82] Bojke L, Claxton K, Bravo-Vergel Y, et al. Eliciting distributions to populate decision analytic models. Value Health. 2010;13(5):557–564. [published Online First: 2010/03/30].

[83] Veen D, Stoel D, Zondervan-Zwijnenburg M, et al. Proposal for a five-step method to elicit expert judgment. Front Psychol. 2017;8:2110. [published Online First: 2017/12/21].

[84] Discounting for public policy: theory and recent evidence on the merits of updating the discount rate: council of economic advisers issue brief; 2017. [cited November 2019]. Available from: https://obamawhitehouse.archives.gov/sites/default/files/page/files/201701_cea_discounting_issue_brief.pdf

[85] Severens JL, Milne RJ. Discounting health outcomes in economic evaluation: the ongoing debate. Value Health. 2004;7(4):397–401. [published Online First: 2004/09/29].

[86] Clay E, Pochopien M, Aballea S, et al. PRM44 - Differentiating discount rates in cost-effectiveness evaluations in the context of gene therapies. Value Health. 2018;21:S363.

[87] Lipscomb J, Weinstein MC, Torrance GW, et al. Time preference. In: Gold M, Siegel J, Russel Leditors. Cost-effectiveness. Heal. Med. New York, NY: Oxford University Press; 1996. p. 214–246.

[88] Cleemput I, Neyt M, Van De Sande S, et al. Belgian guidelines for economic evaluations and budget impact analyses: second edition. KCE reports. Brussels: Belgian Health Care Knowledge Centre; 2012. [cited May 2018]. Available from: https://kce.fgov.be/sites/default/files/page_documents/KCE_183C_economic_evaluations_second_edition.pdf

[89] National Health Care Institute. Guideline for economic evaluations in healthcare. Diemen; 2016. [cited May 2018]. Available from: https://english.zorginstituutnederland.nl/publications/reports/2016/06/16/guideline-foreconomic-evaluations-in-healthcare

[90] Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. N Engl J Med. 1977;296(13):716–721. [published Online First: 1977/03/31].

[91] Keeler EB, Cretin S. Discounting of life-saving and other nonmonetary effects. Manage Sci. 1983;29(3):300–306.

[92] Claxton K, Paulden M, Gravelle H, et al. Discounting and decision making in the economic evaluation of health-care technologies. Health Econ. 2011;20(1):2–15. [published Online First: 2010/12/15].

[93] Arrow K, Lind RC. Uncertainty and the evaluation of public investment decisions. Am Econ Rev. 1970;60(3):364–378.

[94] Bloomberg. Government bond rates overview; 2020. [cited 2020 June 10]. Available from: https://www.bloomberg.com/markets/rates-bonds/government-bonds/us

[95] Klock RM, Brouwer WB, Annemans LJ, et al. Towards a healthier discount procedure. Expert Rev Pharmacoecon Outcomes Res. 2005;5(1):59–63. [published Online First: 2005/02/01].

[96] Brouwer WB, Niessen LW, Postma MJ, et al. Need for differential discounting of costs and health effects in cost effectiveness analyses. BMJ. 2005;331(7514):446–448. [published Online First: 2005/08/20].

[97] Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. Health Econ. 2001;10(7):587–599. [published Online First: 2001/12/18].

[98] Ultsch B, Damm O, Beutels P, et al. Methods for health economic evaluation of vaccines and immunization decision frameworks: a consensus framework from a European vaccine economics community. PharmacoEconomics. 2016;34(3):227–244. [published Online First: 2015/10/20].

[99] WHO guide for standardization of economic evaluations of immunization programmes; 2019.